Davide Botting: The Collectivity of Blaming ......................... 1

Odai Al Zoubi: Austin on Literal Meaning ............................. 41

Nenad Natan Popovic: New Experiments On Psychological Ego-

ism .................................................................................. 65

Charlotte Unruh: Present Rights for Future Generations ........ 77

Johannes Friedl: Erkenntnisbasis und Intersubjektivität. Moritz
Schlick und Béla Juhos über Konstatierungen ...................... 93

Enno Fischer: Conference Report: SOPhiA 2016 ................. 121
EDITORIAL

KRITERION – Journal of Philosophy is a forum for contributions in any field of analytic philosophy. We welcome submissions of previously unpublished papers, not under consideration for publication anywhere else. Submissions are reviewed in double-blind peer review mode. Contributions should meet the following conditions:

1. The content must be philosophical.
2. The language must be intelligible to a broader readership.
3. The contribution must contain a traceable argumentation.

The length should be between 4000 and 8000 words. Only contributions in English (preferred) and German are accepted.

IMPRESSUM

EDITORS-IN-CHIEF: Christian J. Feldbacher-Escamilla, Alexander Gebharter
EDITORIAL BOARD: Albert J. J. Anglberger, Laurenz Hudetz, Pascale Lötscher, Christian Wallmann
ADDRESS: Franziskanergasse 1, 5020 Salzburg, Austria.
E-MAIL: editor@kriterion-journal-of-philosophy.org
WEB: http://www.kriterion-journal-of-philosophy.org
INDEXING: KRITERION – Journal of Philosophy is indexed and abstracted by The Philosopher’s Index and EBSCOhost Humanities Source. Information about the journal’s ranking is available at SJR. The journal was also approved of satisfying the ERIH (European Reference Index for the Humanities) criteria: ERIH PLUS.
COPYRIGHT: The copyright remains with the authors.
ISSN: 1019-8288
Abstract

In this paper I want to argue that acts of blame are performed by collectives, and not by any collective but only by collectives that satisfy certain conditions – broadly those that, by collectivizing reason, can be held to be autonomous subjects to which it makes sense to attribute attitudes, including participant reactive attitudes such as resentment. The actors involved must also be related to the collective in particular ways in order to hold and be held responsible, but they need not have the same attitudes as the collective.

This implies both that our attributions of moral responsibility to an agent does not depend only on facts about the agent but also on facts about us, and that the reasons for which we hold responsible and by which we justify our moral practices, and which the agent can likewise appeal to in an attempt to avoid being held responsible, depend on the collective involved and its substantive ethical precepts. The answer to whether an agent is responsible, or is an appropriate object of a reactive attitude, is not to be settled purely by philosophical analysis, and even the kind of naturalistic description of the pattern of our reactive attitudes that [22] suggests in “Freedom and Resentment” will only go part of the way – the excuses and exemptions are things we might say in a dialogue where reasons are exchanged, but each collective has some leeway over what substantive reasons count as grounds of an excuse or exemption.

Keywords: responsibility, blame, reactive attitudes, collective intentionality, group speech-acts

1 Introduction

Peter steals from Paul, is caught, and the judge sentences Peter to a year’s imprisonment. The judge has blamed Peter, punished him, and
we can say that his doing this counts as the community blaming and punishing Peter, for the judge is authorized to act on behalf of the community in these matters. We can agree to this even if, motivated perhaps by hard determinist intuitions, we do not think that Peter is morally responsible and consequently not subject to a specifically moral form of blame. I will be concerned in this paper with genuine cases of moral responsibility, but to do that I need to explain where exactly the moral dimension enters into our moral practices and what it is needed for, and to do that I need to describe the cases where it is absent.

An incompatibilist will believe that Peter is not morally responsible and that our blaming Peter is not morally justified if his actions were determined. A hard determinist will believe further that Peter’s actions are determined, and conclude that Peter is not morally responsible, and in fact that all moral concepts are empty, having no instances. It is not impossible (although it would be unusual) for such an incompatibilist to accept in part the kind of compatibilist view urged by Smart and Schlick that justifies our practices of blame and punishment instrumentally on the grounds of their deterrent effect and general usefulness. But the incompatibilist, even one who accepts this instrumental kind of justification to be valid, can still insist that this is not generally how we justify, because generally we do not conceive of blame fundamentally as regulation and management of social behaviour but take it to have a moral weight over and above its usefulness, and to require a correspondingly moral justification; he is likely to say also that if we look at what we offer in the way of justifying our moral practices, as a matter of empirical fact what we will find is that our justifications do concern themselves with moral considerations and that considerations of utility would be seen to be inappropriate or changing the subject. If we blame without conceiving of our blame in this particular way, then this is not to count as a case of moral blame. The kind of compatibilist who takes instrumental justification to be all there is cannot accept any genuine cases of moral blame, and attempts to morally blame would involve such a blamer in a pragmatic inconsistency. Not all compatibilists are like this, though, and some would agree with the incompatibilist that moral responsibility requires free will but provide an account of free will intended to be consistent with determinism without adjuring any conceptual revision upon us, yet still strong enough to ground those concepts, e.g., of moral responsibility. If this succeeds, both some compatibilists and incompatibilists can be committed to there being such a dimension, while both hard determinists and Schlickian compatibilists are committed to there
not being such a dimension.

I have explained how the moral dimension enters into our moral practices and what it is needed for: to be engaged in and see oneself as engaged in a moral practice commits one to a conception of blame where it has moral weight beyond its usefulness, and this is reflected in our second-order practices of justifying those first-order practices and our moral discourse in general, which involves the exchange of moral reasons and not just instrumental reasons. One could not knowingly reject this conception and still see oneself as doing the same kind of thing; it is possible that we may still do the same things, but we could not conceive of what we are doing in the same way without pragmatic inconsistency.

I am going to labour this important point, because it is important to the kind of inquiry I am engaged in: the guiding intuition is that we take our practices of blame to have moral weight, and even if we are wrong about this, we would be acting inconsistently if we attempted to blame – conceiving blame here as we customarily conceive of it – while not believing our blame to be morally justifiable and for the person we blame not to deserve that blame. When we attempt moral blame, a fortiori we are committed to a certain conception of moral blame and of moral responsibility, whether we are aware of it, or are able to formulate or articulate it, or not. This is not a conceptual inconsistency and I am not engaged in a conceptual inquiry or analysis, except incidentally. The question is not whether determinism and moral responsibility are compatible concepts, but the pragmatic consistency of holding two views simultaneously.

However, although I will come back to the problem of free will later in the paper, it is not my primary concern here, which is rather to answer the question: seeing that moral practices are acts carried out in and by communities, what must the community be like in order to be justified in carrying out these practices? What does the conception of responsibility presupposed in these practices require of its practitioners so that they may practice without pragmatic inconsistency or performative contradiction? Let us pose the question more specifically: how does the judge’s authority have moral weight, and how might it fail to carry this weight?

If social regulation were all there was to it, there would be no moral weight and there would be nothing more to the matter than the legal powers that have been distributed into various authorities — the judge’s power to blame and punish, and represent the community as he does so, is given him by the legal system. Indeed, even when social regulation is
not all there is to it authority of this kind has this kind of backing. We must distinguish then between the authority that is invested in the judge in the form of his legal powers and responsibilities, and the distinctively moral authority he has in virtue of the community being a certain way such that the community has the authority to issue moral demands to its members, which is to say, in virtue of the community being characterizable as a *moral community*. In principle, it is quite possible to object that a judge lacks the moral authority to pass judgment in a particular case without saying that he is in any way abusing his legal powers when he does so.

In a sense Paul also acts on behalf of the community, for although the reactive attitude he holds towards Peter is a participant reactive attitude, Paul’s reasons are agent-neutral though relative to the moral community whose norms Peter is deemed to have violated. That is to say, Paul believes that the reasons he has would generalise to cover any relevantly similar situation; this is just part of the phenomenology of those attitudes we count as moral. So another question is: how must Paul be related to Peter, and to the community? And as for Peter, how must he be related to this community in order to be held morally accountable to it? To what extent does his personal belief in the community’s values (or lack thereof) or enculturation in the values of a different community, excuse him?

I want to show what kind of obligations these three persons have to each other and to the community, and gradually refine the conditions that the community must satisfy for these obligations to have force, that is to say, to define the kind of collective that the community must be in order to be a moral community. According to reactive attitudes theory, when moral blame (in the form of holding – but not necessarily expressing – a reactive attitude) is appropriate the agent blamed is morally responsible, and the blame gets its force and its backwards-looking character because it expresses reactive attitudes. This is largely true, except that I argue that blame is a collective act and it is a collective reactive attitude that the act expresses; the community must be constituted in such a way that it makes sense to treat it as an intentional subject and attribute resentment to it.

One of my main claims is that, because of the way that the attitudes we can attribute to collectives supervene on those of the members of the collective, there can be cases where the agent is morally responsible and yet is not blamed, for the community can be said to blame the agent – or perhaps we might say that it has an obligation or commitment to
blame the agent – even when no member of the community blames the
agent, or feels a reactive attitude towards him, or judges that a reactive
attitude would be an apt response. I think that it is because there is an
air of paradox in this that we have many of the intuitions that we do,
especially when we worry over whether we have the moral authority or
standing to hold an agent responsible; many of those intuitions appear in
a different light when the blamer takes the viewpoint of the community
and takes his blaming behaviour to be, in some degree at least, on the
community’s behalf.

2 The Judge

2.1 The basis of moral authority

One thing that must be true is that the judge has the standing to pass
sentence, because he is authorized to do so by the community. It will
be his job to adjudicate on questions of law, that is to say, to ensure
that the rules of procedure for conducting a trial of this kind have been
observed. When this is so, the verdict of the court (whether the judge’s
in the case of a non-jury trial or the jury’s in the case of a jury trial) is
binding. He blames Peter, and punishes him; when one with authority
blames, this counts as the community blaming.

Note, though, that the judge does not necessarily have any feeling of
moral indignation towards Peter, or even necessarily believe that Peter
has done something morally wrong. He may believe that the crime Peter
committed should not be considered a crime, or as a less serious crime;
it may even be something that the judge himself has done on occasion.
Rather, Peter’s guilt comes down to this: when presented with the op-
portunity to give reasons that would have exonerated him or mitigated
punishment (assuming procedure is correctly followed), he failed to do
so, either by giving no reasons at all or by giving reasons that were not
deemed acceptable. The community was obliged to give Peter that op-
portunity, and Peter was obliged to give reasons. Because Peter lacked
a winning strategy in the dialogical exchange of reasons, he was found
guilty. Had Peter not been presented with the opportunity to present
reasons, then the verdict would not be morally justified, binding or have
any moral weight.

How does the judge get his moral authority? It is implausible that
one need to satisfy such stringent demands as for legal authority in or-
der to have the moral authority to blame. Paul, for instance, will blame
Peter by feeling resentment towards him, and these feelings will gener-
ally, at least, be appropriate. [22, 7-10] describes various excuses and exemptions that would be considered appropriate reasons for modifying reactive attitudes, but in the absence of these it seems that Paul, at least, appropriately blames Peter and has the standing to do so. It seems also that it would be appropriate for other members of the community to feel indignation towards Peter.³

Standing and moral authority, then, can have a much more informal basis than legal authority. On Tuomela’s account, there must obtain the right social and normative circumstances, as given by formal and informal rules, underpinned by “general constitutive rules concerning the purposes and functions of the collective” and “proper social norms specifying his social roles” [23, 8]. We can say that when one with authority blames (although not only then), the community has a reactive attitude, and the blame expresses this attitude.

2.2 Supervenience, dialogical commitments, and the autonomy of collectives

How does it make sense to attribute the community with a reactive attitude? Doesn’t this amount to treating the community as an autonomous, intentional subject? Yes, it does, but the autonomy in question is of a weak kind because the attitudes of the collective supervene on – and can be given a (possibly very complex) analysis in terms of – the attitudes of the members of the collective. The analysis involved here may not be of a straightforward kind, such as every member of the collective believing \( p \) making it the case that the collective believes \( p \). On the contrary, it is quite possible for the collective to believe \( \text{not-}p \) when every member of the collective believes \( p \), or for the collective to believe \( p \) when every member of the collective believes \( \text{not-}p \).⁴

We have already seen this in the case of the judge. The judge, I said, ensures that procedure is correctly followed. Think of this procedure like a rule of inference. Given that certain premises are agreed to hold (i.e., that the case presented reaches a certain stipulated standard of proof), and given that the rule of inference has been agreed to be normative, the result is obtained by applying the rule of inference to the premises; this is referred to as meeting the burden of proof. This is a premise-based procedure, and there are cases where the result of following this procedure conflicts with what the members of the collective (say, the jury) believe. Every member of the jury may believe that the accused is guilty yet still rightly consider “not guilty” to be the correct verdict. This is because
the collective chooses to “collectivize reason” rather than suffer the radical discontinuities that can occur, for instance, in cases of discursive dilemmas. Instead of supervening on members’ beliefs about whether Peter is guilty, the group’s verdict about whether Peter is guilty is based on the members’ beliefs about things like whether Peter performed the act, knew what he was doing, etc., and that if these conditions are satisfied Peter should be found guilty. If they accept this last as a valid rule, then the procedure to ensure group rationality is to take some function of their beliefs that these conditions are satisfied depending on the standard of proof required, e.g., a majority or unanimous vote, and to apply the rule \[12, 95-99\]. \[19\] calls this a premise-based procedure, because only the individuals’ judgments of the premises are taken into account, and the conclusion calculated by a rule of inference they accept. This is precisely the case of the judge described above. Furthermore, because the judge and/or jury act on behalf of the moral community as a whole in virtue of the position they occupy (due to Tuomela’s conditions being satisfied), I would say that this group can be attributed with having the belief that Peter is guilty, and even if it so happens that no individual actually believes that Peter is guilty we might say that they accept that Peter is guilty.

The whole process can be usefully represented in the form of a dialogue in which the premises and rule have been agreed on at the start, and then simply applied by following the rules of dialogue. The contributors to the dialogue, once it has been shown that some proposition \(g\) follows from their other commitments, must likewise be committed to \(g\). There is a very close analogy here between an acceptance and a commitment in dialogue. So, if Peter is unable to provide an excuse or exemption in the dialogical exchange of reasons, and the rule that in the absence of an excuse or exemption the agent is an apt subject for reactive attitudes and for being held responsible, the rules of dialogue generate a commitment to hold the agent responsible. This is a commitment for the community itself and for all participants in the dialogue, although it may be enacted only by individuals in the community with the authority to do so.

Let me explain what I mean by a commitment in dialogue. Suppose that I believe that I am less than six foot tall, to express which I utter the words “I am less than six foot tall.” When I do so I am performing the speech-act of assertion, since I am satisfying the conditions for performing this speech-act, namely (in part) that I am sincere, or in other words, that I believe what I say. If I do not believe that I am less than
six feet tall, then I cannot perform the assertion that I am less than six feet tall, and if I try to do so simply by saying “I assert that I am less than six feet tall” then the attempted speech-act *misfires* because some condition of its performance is not satisfied.

A commitment is not a mental state but more like a contractual obligation to an audience, who can justifiably act in expectation of a participant in the dialogue behaving consistently with their commitments (or at least the ones that the dialogue have made them aware of) and can justifiably sanction participants who do not. This is quite reasonable because it can be inferred from the fact that an assertion has been performed that whoever asserts believes what they say. Of course, an audience may not know whether the speech-act actually is a genuine assertion or a misfire, but I become committed to the propositional content of whatever I utter in an assertive mode. It is no excuse to say, after the fact, “I was not sincere when I said that, so you cannot hold me to it,” unless there was an attempt at the time of utterance to cancel the implicature that I should be understood as uttering an assertion and thus as committing myself to a truth-claim regarding the propositional content.

If caught out in an inconsistency then the rules dictate that I must withdraw one of the inconsistent commitments; in short, I cannot stubbornly continue to contradict myself. Why not? Conceding that the rules model a rationality constraint, why allow myself to be held to rationality constraints? Even if it is granted that I am acting irrationally and unreasonably, does it follow that I am acting immorally? Similarly, Pettit has shown that following a premise-based procedure is rational and that it is the way for collectivized reason to avoid inconsistency, but has not shown that it has any moral force; one can be irrational without being immoral, and intersubjective agreement on its own does not do the job, because one can break such agreements without necessarily incurring moral criticism.

It is for this reason that the authority requirement must be satisfied as well as the procedure to be followed for whatever issues from it to be group-binding — to have normative force in such a way that a group-member who behaves in a way incompatible with an accepted proposition may rightly be morally sanctioned by the group, though he is not morally required actually to believe that proposition. That is to repeat that 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.\textsuperscript{6} In speech-act terms, the position-holder’s public acknowledgment of the proposition misfires in much the same way that a baptism misfires if not performed by an ordained minister.

In summary, when reasons are exchanged in dialogue by, typically, making assertions, they incur commitments. It is dictated by the rules of the dialogue what reasons can be appealed to, but I propose that in the context of moral responsibility it is a constraint on these reasons that they fit into Strawson’s framework of excuses and exemptions. It is also dictated by the rules that each partner in the dialogue gets the chance to put forward reasons and to challenge the reasons put forward by the other. Reasons are conclusively established by showing that they follow from commitments already incurred, since one cannot incur a commitment that contradicts a prior commitment on pain of inconsistency. If the agent has failed to give reasons then he loses the dialogue, which is ended, I propose, by a speech-act of blame. Even if the speech-act is not actually performed, though, this does not mean that the agent is not morally responsible, there still being a commitment to blame generated by the rules of dialogue.

When there is a commitment to blame, we can say that resentment can be attributed to the community; this feeling of resentment supervenes on the appropriate members of the collective having certain beliefs about the premises and the inference-rule (i.e., that the agent’s reasons are unsatisfactory), and correctly applying the rule to those beliefs, or in dialogical terms agreement about certain starting-points and argumentation schemes that may be used in the dialogue, and the rules of dialogue themselves being correctly followed. In other words, the resentment results from a premise-based procedure and occurs only in collectives where reason, or at least, reasoning about moral responsibility, has been collectivized.

This speech-act of blame both expresses the community’s belief that the agent is morally responsible and also the resentment that can be attributed to the community because it supervenes on the attitudes whose dynamic, dialogical interplay has resulted in this commitment. I will call this a reactive commitment. Note that the attitudes in the supervenience base need not be reactive attitudes, but note also that it is the same rules that regulate those reactive attitudes that regulate also the conduct of
the participants in the dialogue, and that they can only ever appeal to
reasons that can be discovered among the natural facts; we may discover
these rules by examining our reactive attitudes, but neither blame itself
nor the commitment to blame depends on those reactive attitudes, and
the reactive attitude the blame expresses is the resentment that I have
just said can be attributed to the community. This can be thought of as
the community having an obligation to blame. 7

I should say some more about this obligation, since it is closely tied
to the collective’s being morally autonomous in a certain (weak) sense.
This kind of autonomy was argued for by [6]. Copp proposes two theses:
a weaker and a stronger. The weaker thesis claims that we can ascribe
moral properties – such as responsibility and obligation – to collectives,
where we cannot ascribe the same properties to any of its members.
In other words, collective C can be attributed moral property M1 even
though no member of C can be attributed moral property M1, but at
minimum some moral property Mn must be attributable to some member
of C and it is in virtue of this (at least in part) that C possesses M1.
In other words, it is possible that a group is responsible for a particular
act but we would not like to say that any of its members are responsible
for that particular act. I have defended this claim in [2] and argue that
this is sufficient to achieve the result that collectives can be considered
as having some moral autonomy, but I do not think that we should
conclude from this that they should be considered as agents in their own
right. Copp effectively does for moral properties what Pettit does for
beliefs, and an analogous supervenience claim holds here also: the moral
properties of the group supervene on the moral and non-moral properties
of its members. The group is weakly autonomous when there must be
at least some moral properties in the supervenience base. If there are
none, this is strong supervenience.

Keeping this in mind with regard to what was said earlier about
collectivised reason; collectivising reason does not amount to the claim
that a collective literally has beliefs. Rather, a collective has a set of
propositional contents (formed in the prescribed way) towards which
the appropriate attitude to be taken by the members of the collective
is the practical attitude of acceptance. Nonetheless, these propositional
contents do depend on and supervene on the beliefs of the members,
just not beliefs towards that particular propositional content, as we have
already seen, but to contents relevant to it, i.e., the premises. To say that
that a collective believes \( p \) is to say that \( p \) is one of those contents, and
the usefulness of attributing a belief to the collective is to emphasize the
fact that it makes sense to say that a collective believes \( p \) even when none of its members do — they only accept \( p \). Moreover, when the collective is a moral community, it makes sense to say that the members of the collective have moral obligations to accept these propositional contents, because Tuomela’s conditions guarantee that these collective beliefs are normative and group-binding and at least sometimes (depending on the belief) this normativity is moral.

This idea extrapolates beyond beliefs. We can attribute moral sentiments, moral obligations and moral responsibility to the moral community that are distinct from any that we can attribute to its members — either because of having different propositional contents or because of having different circumstances in which they might be inappropriate — that nevertheless depend on those we can attribute to its members. There are two sets of circumstances in which it might be inappropriate for someone to blame: either they lack the moral authority as we have already seen, or; they have the authority and the moral obligation to blame, but that obligation is overridden either by another obligation or by a reason undermining their moral responsibility for not blaming, as we will see later. I want to be able to say in such cases that the moral community still has this moral obligation and can be morally responsible for not complying with it even when no member of the community can appropriately be held to it, and it will be shown in the next section how Copp argues for this result and that this amounts to the collective in question being autonomous. A moral community must be autonomous in this weak sense.

Collectives must be autonomous enough to be the bearers of intentional and moral properties such as belief and resentment on the one hand and moral obligation and responsibility on the other, without being autonomous so strongly that belief in some propositional content could be attributed to the collective without any beliefs at all (e.g., about its premises) being attributed to its members, or moral obligation to blame could be attributed to the collective without any moral obligations at all being attributed to its members. Copp’s stronger thesis claims that we can ascribe moral properties to collectives where we cannot ascribe any moral properties at all to any of its members such that the moral property ascribed to the collective is constituted by or supervenes on these moral properties. This result is held to be shown by discursive dilemmas, but I hold that these dilemmas show only that the relation between the properties of the collective and its members can be exceedingly complex, and reject this stronger thesis. In other words, I hold that
there must always be something of the same kind as what is attributed to the collective in the supervenience base; if there were not I would call this autonomy strong, and I am sceptical whether collective autonomy as strong as this is possible. Fortunately, I do not need it, and no more will be said about this stronger thesis.

[6, 370] employs a distinction between all things considered obligations and responsibilities and pro tanto obligations and responsibilities. A pro tanto obligation is an intrinsically but defeasibly motivating reason for action with moral force. Thus an agent may have a pro tanto obligation to A but may not have an all things considered obligation to A, e.g., because of a conflict of duties out of which the duty to perform A is less pressing than some other incompatible duty. Similarly, the agent may be pro tanto responsible for A but not all things considered responsible for A. Whilst it is intuitively obvious what it means to have an all things considered or pro tanto obligation, what does it mean to be all things considered or pro tanto responsible? [6, 370] defines it as follows:

I stipulate that an agent is responsible ‘pro tanto’ for having done A just in case there is a moral basis for holding the agent to be deserving of a negative or positive moral response of some kind for having done A, where, in the absence of a countervailing considerations such as an excuse or justification, this reason would be sufficient to make it the case that the agent deserves the response all things considered.

This definition is clearly utilizing a reactive attitude theory of responsibility. An all things considered responsibility derives its categorical force because it occurs after all relevant considerations have been taken into account.

Let us return now to the speech-act of blame that in ideal circumstances terminates the dialogue. If the appropriate position-holder does not blame, despite there being a reactive commitment or equivalently an obligation to blame, then this is something for which the position-holder can validly be held to account, and for which he might be held morally responsible. This, I think, is what happens in many cases: the one whose task it is to blame does not blame for various good or bad – but easily understandable – reasons that generate the intuition that blame would be inappropriate, or that it would be inappropriate for us, or for some individual, to blame. Such an individual would try to justify his not blaming on the grounds of believing that blaming would have involved a performative contradiction, such as that he did not have the standing to
blame in this particular instance. Perhaps for some individuals the simple fact that he does not experience a reactive attitude would be a reason not to blame, as indeed would be the case were experiencing a reactive attitude a condition of satisfaction on the blame as belief in the truth of the proposition is a condition of satisfaction of asserting it, but it is not such a condition. It is that such a reactive attitude be attributed to the community that is a condition of satisfaction and not the individual’s. The individual errs, then, when he takes his own reactive attitude or lack thereof to affect his standing to blame or justified-ness in blaming.

To repeat: absence of a reactive attitude is not a good reason for withholding blame, for not holding an agent responsible, for it is not a requirement of blame that the one who blames experiences a reactive attitude: this is reflected in my model by the fact that the speech-act of blame is performed by the community, and it is the beliefs and attitudes of (i.e., that can be attributed to) the community that are involved in the conditions of satisfaction of the speech-act of blame. In other words, although it is an individual who blames or carries out a speech-act, it is not the individual, but the collective on whose behalf he acts, that needs to have the right attitudes for a successful blaming and that are referred to in the speech-act’s conditions of satisfaction.

However, there are also cases where it appears that the position-holder’s pro tanto obligation to blame could in principle be over-ridden by another pro tanto obligation, without apparently over-riding the community’s pro tanto obligation. This is because the individual involved may have a pro tanto obligation that applies to him personally (e.g., caring for a sick child) that are not plausibly obligations also of the collective to which the individual belongs, for which reason it defeats the individual’s pro tanto obligation to blame without defeating the collective’s pro tanto obligation. This, I think, would be a valid reason not to blame but, because the community’s obligation to blame is still in force and, being undefeated, it is in consequence the community’s all things considered obligation, we still have the result that the agent is morally responsible.

There is, then, a division of labour in my model: the conditions of satisfaction of the speech-act of blame, and the conditions that the collective must satisfy in order to perform that speech-act. So far I have been focussing on the latter.
2.3 The collective perspective and the individual perspective

That the latter is required at all might seem to present a problem for identifying the agent’s being morally responsible with the appropriateness of moral blame. Only moral communities — collectives satisfying certain conditions — can blame, and only when it is someone who is authorized to blame by the community who blames is the blame appropriate and not a performative contradiction of some kind. This implies that whether an agent is morally responsible does not depend only on facts about him but on facts about the collective and about the agent’s relationship to it, on whether the collective satisfies the right conditions, and on whether the one who blames satisfies the right conditions.

Smith tries to use this kind of intuition to drive a wedge between an agent’s being responsible and our appropriately holding them responsible: we need a standing in order to hold someone responsible that we would not take to undermine responsibility itself, that is to say, cases where although we would agree that an agent is responsible we would not be willing to hold them responsible or even judge that it would be appropriate for us to do so. She gives cases like the following: feeling resentment about how one person treats another in their relationship while not believing that it is your place to “interfere” or take up an attitude of blame towards him, or not blaming behavior that you believe to be reprehensible out of a feeling that “People who live in glass houses should not cast stones”; e.g., because you are guilty of comparable misdemeanours [20, 478-480]. Thus, “holding responsible” is not just a question of determining what psychological states the actor is in but, putatively, being responsible does depend only on the psychological states of the actor, in part on her motives/reasons for acting. This seems to offer the biggest obstacle to the identification of being responsible and holding responsible.

I think that this is actually a positive feature of this account, although it does imply a certain, irreducible degree of relativism: in an extreme case, it is possible that what is morally approved of in one culture would be morally condemned in another, and that an agent would be morally responsible in one group for some action but not in another.

Is it true, as Smith objects, that we sometimes do not hold someone responsible because we do not have the standing to hold them responsible, but that we nevertheless consider the person in question to be responsible? Does it follow that being responsible and being held responsible are not the same? I concede that if we take the individual perspective, a person may believe that they do not have the standing
and decide therefore that it would be inappropriate for them to blame, and foreswear blaming. But, taking the perspective of the moral community, as long as you satisfy the comparatively weak authority conditions for blaming behaviour then this belief that you lack the standing is false and you do hold the person responsible, even perhaps if you think you do not, and I have argued also that the moral community does blame the person in question.

Similarly with hypocrisy: is it true that I lack the standing to blame if I have been guilty of the same kind of violation? Is non-hypocrisy a speech-act condition of blaming? Furthermore, is it actually inappropriate to hold someone else to an expectation or demand that one would not hold oneself to in the same circumstances? Does a blamer need to be consistent and, if so, why? Is hypocritical blame non-blame?

Wallace’s answer is, briefly, that when we blame hypocritically we are implicitly taking ourselves to be of a different moral standing to the one we blame, and by doing so fail to respect our equality as persons ([24]). I find this less than obvious and, in fact, I do not think that a blamer must be consistent over time. In fact, I believe that when I blame an agent for an action I implicitly blame all relevantly similar agents performing relevantly similar actions. If I myself have performed such actions then I fall within the scope of my own blame, and I may indeed realise this and deeply regret my past transgressions and see them for what they were and see the excuses and justifications I may have appealed to at the time to be the self-deceptions they were. The ‘inconsistency’ involved here seems to be nothing more or less than conscience.

Thus, there is no moral fault in hypocritical blame — it is often the first step on the road to moral realisation and non-hypocritical blame. This is not to say that being hypocritical is not a morally reprehensible quality in and of itself, but there need not be anything defective in the blame itself or unwarranted in giving the blame its usual force, especially since this force ultimately comes from the moral community itself. “People who live in glass houses should not cast stones” is not a moral principle, and the intuition that hypocrisy defeats the obligation to blame comes ultimately from taking the individual perspective rather than that of the collective.

This erroneous charge of hypocrisy comes up in another way. Remember that the commitment to blame is always attached to the particular dialogue in which it occurs. We can see this in our own culture over an extended period of time: accepted practices of those cultures (e.g., slavery) are deeply objectionable to the cultures that succeeded them,
though in some sense it is the same culture. Are we then being hypocritical when we condemn those cultures? Granted, I think there is reason for saying that they are not responsible, but this, I think, is because it makes little sense to think of a society that has passed away as being in a dialogue with us now, or of our obliging the society to justify itself, or of it being obliged to justify itself to us. The intuition that it would be hypocritical and that it is specifically this that undermines our holding them responsible – though widespread and powerful enough to explain our feelings of inappropriateness – has nothing to do with it.\(^9\) We are not prevented from holding those involved in those practices responsible in some sense, and even, arguably, bearing some of that responsibility ourselves, as the successors of that culture.\(^10\) We have already seen how a member of a collective can be obliged to accept something even if he does not agree with it, and in consequence he is pro tanto responsible for it. Cultural ignorance is not, itself, a case of exemption, but it may be that in some blame societies it is accepted as grounds for an exemption, or at least a mitigating circumstance in the punishment phase. All that matters is a participatory intention, and while this intention should be strategically responsive to cultural changes, it is by no means obliged to bring those changes about; on the contrary, the obligation is always in favour of the status quo, just as it is the pro tanto obligation of the judge to follow procedure irrespective of whether he believes the procedure to be correct, for it is only because the procedures are followed correctly and seen to follow correctly that what issues from them is binding on all members of the community.

*Participatory intentions* and *strategic responsiveness* are technical terms introduced by \([10]\). Rather than introduce Kutz’s account here, I will note this point about strategic responsiveness follows from there being a pro tanto obligation to follow a procedure – or, in this case, to conform with a customary practice – and that this obligation will often win out against the pro tanto obligation of the judge to change the procedure and to not conform to the practice out of a conviction that the procedure is defective. Generally the first priority is to follow procedure, because it is precisely by following procedure that reason is collectivized and attitudes can be attributed to the collective; it is by following procedure that the position-holders act on behalf of the collective, and if procedure is not followed then one exceeds one’s authority as a position-holder and whatever speech-acts are performed by the individual will not be binding on the collective. The validity of the outcome – what makes the outcomes of the individual deliberations qualify as the decision of the collective –
is following those procedures, even if from a wider point of view another procedure would be fairer, and even if the individuals know this. If, however, a better procedure is introduced then the individuals should be responsive to that fact. So, although there is, I think, a pro tanto obligation to change a procedure believed to be incorrect, I do not think that this obligation necessarily wins out against the pro tanto obligation to follow procedure. The individuals must be strategically responsive, but not necessarily pro-active. This is just another case where an individual finds himself obliged to accept qua his social role and position something he personally does not believe or endorse.

Kutz’s account attempts to characterize a kind of collective that is more weakly bound together than those that the analysis of Tuomela is trying to capture. I do not intend to weaken thereby the conditions that need to be satisfied for a community to be the kind of collective that can blame; Tuomela’s account described how a collective must be in order to blame, whereas Kutz’s account describes how a collective must be in order to be blamed, for where the cultural practice is morally reprehensible it is the collective whose practice it is that is blamed first and its members second and incidentally.11 Possibly there is an asymmetry then in the strength of the conditions collectives must satisfy to occupy their respective positions in the blame relation. More likely, though, both collectives must satisfy the stronger conditions, although it is the weaker conditions that account for the kind of responsibility that an ordinary member of a collective may have simply in virtue of being a member of a collective, irrespective of what position he occupies within the collective. If a collective act or practice is morally reprehensible, then this reprehension is shared by every member of the collective no matter what their personal feelings on the matter are or how they conceive of what they are doing when engaged in the practice. It is what they accept that counts.

This is not to say, of course, that if the collective is morally responsible then its members must be morally responsible — we have already seen that this is not true because of the possible existence of pro tanto obligations possessed by individuals but not by the collective they belong to. For the same reasons a collective may hold an agent morally responsible, without any of its members holding an agent morally responsible. Unlike the standing and hypocrisy cases, where I argued that the intuitions generated, though understandable, were erroneous, here I think the blamer is right not to blame. It does not follow, of course, that there is no longer a pro tanto obligation for the individual to blame,
still less that there is no longer a pro tanto or even all things considered obligation for the collective to blame. Because there is an all things considered obligation for the collective to blame the agent is deemed to be held responsible by the collective and to be responsible. This is not a case where the agent is responsible but it is inappropriate to hold them responsible. On the account I have been proposing, Smith is wrong to think that there could be such cases.

Sometimes no sanctions are actually taken against the person either because you consider your obligation to sanction to be defeated by other factors or because you do not personally share the reactive attitude and, now taking the individual perspective, are not prepared to act upon it. Also, one is more inclined to take the individual perspective towards someone with whom you have a personal relationship than otherwise. Consequently, you may take a different attitude towards cases of negligence, or towards cases of a husband being boorish towards his wife, depending on whether the parties involved are friends or strangers. This does not mean that the group does not hold them responsible or that your reasoning does not result in blameworthiness.

On the other hand, it does not necessarily mean that the group does hold them responsible either. The possibility is left open that some groups may accept as reasons what others may not; an agent may have a winning strategy in one dialogue that he does not have in another. This means that whether an agent is responsible cannot be determined solely by facts about the agent.

2.4 Accountability and responsibility

We can summarize what we have said so far by saying that to be responsible and an apt object of reactive attitudes an agent must first of all be morally accountable. The idea of accountability is that if an agent is accountable, then there is a prima facie case against the agent that the agent is obliged to answer. It does not say anything about whether this case can be answered, and in many cases it can. Here is the definition of accountability as such:

(Accountability) An agent A is accountable for X if there is a group G such that

i) G is autonomous towards blaming A for X;

ii) A is so related to G that A is obliged to justify X-ing to G;
iii) G can carry out sanctions against A.

I will now gloss this.

The notion of autonomy in (i) is the weak kind I defended earlier that allows us to treat G for certain purposes as if it were a subject. In particular, it can blame, which is conceived on the analogue of a speech-act. Nor is (ii) a very strong condition; being a member of a group certainly does not require endorsing all its values, and the values that she may appeal to in the course of justifying her actions need not be ones that she personally endorses, and she is acting both morally and rationally when she does this. The condition (iii) recognizes that if a group cannot carry out any of the sanctions that it entertains then dialogical commitments have no force. If (iii) is not satisfied then G can only play at blame; it may still perhaps be counted as carrying out an evaluation of A, but this has no force. There must at least be the possibility that the resulting commitment can be expressed and externalized in some act of blaming, whether it actually is or not.

Note that this is a definition of accountability as such, and not specifically moral accountability. For moral accountability I argued above that Tuomela’s authority condition must also be satisfied in order to make the attitudes that can be attributed to the collective (i.e., the group-beliefs) normatively binding on the collective — these are not only attitudes that the group has but that it ought to have and in contradiction to which members of the group ought not to act. When this is so the blame, obligations and (implicitly) commitments involved in the definition can be characterized more or less equivalently as moral blame, moral (pro tanto) obligations and reactive commitments. In such a circumstance I will say that A is morally accountable to or responsible to G. Here is the definition of being responsible:

\[
\text{(RESPONSIBILITY) An agent is considered to be morally responsible if s/he can be morally praised or blamed by a group towards which the agent is morally accountable.}
\]

Praise and blame are success terms. To say that an agent has been morally blamed then implies that the one doing the blaming would satisfy the conditions on the speech-act of moral blame.

### 2.5 Blame and free will

I noted earlier that there was a division of labour in my model: the conditions of satisfaction of the speech-act of blame, and the conditions that
the collective must satisfy in order to perform that speech-act. The conditions that the collective must satisfy in order to perform that speech-act are given in the definitions above. It remains to give the conditions of satisfaction of the speech-act of blame:

(BLAME) A group G ‘blames’ an agent A if and only if:

i) A is in dialogue with G, and the rules of the dialogue allow for A to put forward reasons, and for those reasons to be challenged. (The rules facilitate and cannot prohibit the free exchange of reasons beyond the fact that it must guard against an inconsistent commitment-set).

ii) A has no winning strategy in the dialogue. (A fails to defend his actions by appeal to commitments shared with the group).

iii) The group believes the actions defended by the agent to be raising a question that is “distinctively moral.”

iv) The group believes its own response (the ascription of blame to the agent) to this question to be also “distinctively moral.”

v) The agent is not believed to be the beneficiary of moral luck (good or bad).

It should be noted that these definitions (praise being defined in a similar manner) allow for an agent to be blamed when there is a dialogue between A and G [condition (i)] whether or not A is obliged to be in the dialogue with G, although the inference from blame to being considered to be responsible only goes through when there is such an obligation, as the definition of responsibility shows. The definition of blame also allows for an agent to be blamed when any of the attitudes listed in BLAME (iii-v) are false; sincerity does not imply truth but only believed truth. What this shows is that when we blame and take our blaming to have moral weight, we are implicitly committed to a conception of responsibility that rules out the morally unlucky as candidates for moral responsibility. This is why responsibility is defined above in terms of an agent being considered to be morally responsible, rather than morally responsible simpliciter. To put it another way, it is an account of responsibility-attribution.

Note also that there is no reference here to what the agent’s actual motivating reasons are; for instance, an agent may be praiseworthy even if he did not do the right thing for ‘the’ right reason, but there must be ‘a’
right reason that he shares with the group that can be appealed to. Nor, to an extent at least, does it matter what the group’s actual motivating reasons for blaming are. Consider the judge again, and suppose that he feels, and acts on, a retributive urge towards Peter, so that it is not the reasons that morally justify his blame that actually motivate it. We can use the account recursively, this time taking the judge as the plaintiff. Can the judge justify to the group he represents the act of blaming Peter? Does the judge have a winning strategy? It seems to me that he does: namely, that the conditions of satisfaction on blame were satisfied, and even should this be a post facto rationalization of the blame it does not make the blame any less blame or Peter any less morally responsible. This is not to say that the judge has not acted reprehensibly in allowing himself to judge in this way. Similarly if the blame is hypocritical, that is to say, the judge would not hold himself to the norms that Peter has violated. It is a false (though understandable) intuition that this undermines his moral authority to blame Peter; although hypocrisy and rashness are reprehensible in themselves, this does not mean that hypocritically or rashly blaming Peter is also morally reprehensible.

By analysing the speech-act of blame we find ourselves constrained by its satisfaction conditions; these conditions are basically criteria for determining who can be blamed without performative contradiction, and indirectly an analysis of the conception of responsibility involved in acts of blame.

The first constraint is that only groups – or individuals taking the perspective of the group – can blame. This may appear unmotivated. Why can praise and blame only be performed by a group? It should be pointed out here that it is often the case that it is simply not in a single person’s power to sanction another, and without the ability to sanction and to carry out threats there is no accountability. However, there are also clearly cases where it is in a single person’s power in so far as the individual gets to execute an action – express a reactive attitude – to which the group have given at least their passive consent. So when I say that only groups can praise and blame it does not mean that an individual literally is unable to perform an act, but that in doing so he acts qua member of the group and on behalf of the group (though possibly on his own behalf as well). Should he literally be unable to perform the act this does not mean that he does not have the corresponding attitude but only that he is unable to express it, at least in such a way that it is analysable as a dialectical move affecting commitments; he is merely blowing off steam. Nor does failing to have the corresponding attitude
prevent in any way his performing the act. So, he can have the attitude and fail to perform the act, and fail to have the attitude but still perform the act. This, as I have said, is because the attitudes involved in the blame are those that are attributed to the group. But he must satisfy the conditions for being a position-holder, since otherwise he does not act on behalf of the group and the speech-act does not count as that of blame because it is not then a speech-act performed by the group. Having the authority to act on behalf of the moral community is thus an external condition on successful performance of the speech-act, and does not in fact rely on what the speaker believes, although the speaker would nevertheless be acting inconsistently if he thought that he did not have the moral authority and yet still attempted to perform this speech-act; what I am saying is that the speech-act may still actually be successful when he is mistaken in this belief.

So we can rephrase the question: why is taking the perspective of the group a constitutive condition of speech-acts of blame and praise? One answer is that it is already there implicitly in the reactive attitudes theory when it talks about cultural practices such as those of retributive justice; these are collective acts, and not mere attitudes.

A better answer is that such acts of blame are considered by their perpetrators to be distinctively moral, and this carries along with it the implication that it is how every relevantly similar agent would act when faced by relevantly similar cases. This is part of the phenomenology of moral reasoning and action and is unaffected should it turn out that relevantly similar cases are treated differently by different groups. In principle, it is even possible for one group to praise what another group would blame, and for a person who belongs to both groups to praise qua member of one of those groups what he blames qua member of the other. It is important to realise that such a person is not being inconsistent or making a performative contradiction, since his actions are relative to the group. This is reflected dialectically by the fact that there is a different dialogue and commitment-set for each group to which the person facing sanction must defend himself, and no inconsistency in a commitment-set is produced by this apparently contradictory behaviour. Similarly, the person facing sanction need not defend himself in the same way to each such group. She must defend herself only by appeal to commitments that she shares with her interlocutor, and it is of course entirely possible that she might share commitments with one group that she does not with the other. She, also, is not behaving inconsistently or making a performative contradiction when she does. On the contrary, considered
dialectically she is arguing precisely how she should. What makes this consequence less palatable is that, according to what I said above, if one group praises her and another blames her, then she is (or, rather, considered to be) both praiseworthy and blameworthy.

However, the bullet may not be so indigestible after all, because there are constraints on the group’s activity if it is not to be involved in a performative contradiction, and I will argue that the central aspect of this is the disqualification of moral luck. I start by assuming that the purpose of moral practices is to discourage the morally reprehensible and that the group accepts the principle of “ought implies can.” Luck, by definition, is not something about which we can do anything, so it cannot be the case that we ought to do anything. Of course, it might be argued that this assumption is false in some cases and there are real cases where bad luck is punished, but where it is then I would say that the practices in question are not distinctively moral and that being distinctively moral was one of the conditions of satisfaction of blame. So, such a group plays at blame, but does not really blame. Blame plays some other social function here: as ritual or bonding exercise.

I have now used the phrase “distinctively moral” several times and it could be asked what I mean by this. I will demonstrate what I mean by using it in an argument. Let’s start with an argument from [14] that is in some ways analogous. Malcolm argues not that determinism is inconsistent with intentional action but, interestingly, that the assertion of determinism is inconsistent with intentional action, because when I use language I would not be relying on its intentional properties for communication but only on its causal properties. But this is a false dilemma, and Malcolm has fallen into the trap of confusing determinism with (mechanistic) reductionism; it is quite possible for deterministic causes to include beliefs, intentions and everything else required for successful performance of an assertion or other illocutionary acts.

However, I do not think that assertion of or belief in determinism is consistent with successful performance of acts analysable as blame. If determinism is true, then ascriptions of blame are just as much determined as everything else. This response is an ancient one that goes back to the Stoics but is rarely countenanced in the modern literature. The Stoics did not actually endorse this response, being stern moralists according to Sorabji who at [21, 87] gives details of its ancestry. Sorabji calls this is a hard determinist response, but I think that this is a little unfair. The point is rather that, if we assume determinism, there is no real moral question of whether our blame is just or not, for whether we
blame or not is decided by some causal route over which we have no control; the whole debate becomes beside the point. My point is that we also are moralists, albeit a little less stern than the Stoics, and take our blame to have a moral weight that it could not have were it determined. We do not take the debate to be beside the point but to be very much on point, and this in itself — whatever side we may take in the debate — commits us to a certain conception of responsibility that is inconsistent with determinism. To put it another way, one side of the debate actually seems to be incoherent with having a debate at all.

Thus the judge, in believing that his judgment has a distinctively moral status, must assume that his judging is free, and that at least he himself has libertarian free will; otherwise, talking of morality’s aims is redundant. If everything can be described in a purely descriptive vocabulary, then questions about morality and the values of personal relationships in a deterministic universe are problematic only in the sense that they cannot arise at all.

For example, incompatibilists often complain that determinism diminishes things, such as human relationships, that have great value to us, and compatibilists in their turn take a great deal of trouble to show that this is not the case. But both the question and the answer presuppose that there is a moral issue at stake, that it matters — it occupies a domain whose problems are not settled by simple description, e.g., of human behaviour, and neither side would consider such a description to be apposite. But if determinism is true then it seems as if to every genuine question a description should serve as an answer. Similarly, in my view, the worry that our moral practices are unjustified if determinism is true misstates the problem, for if determinism is true no questions about its moral status can seriously arise.

There is also a methodological implication here. Sometimes empirical psychologists present various scenarios and ask for some kind of moral judgment. When the scenario is a deterministic one and aimed at assessing whether folk intuitions about moral responsibility, to take an example, are compatibilist or incompatibilist, this question is improper, in my view, unless, perhaps, “not applicable” is provided as a candidate answer. It should not be assumed that this is a genuine question, or that “Yes” and “No” exhaust the possible answers; they are not even relevant answers if the question involves a false presupposition.

Scenarios described in physicalist language are simply not the kind of things about which a group is asked to make moral judgments, or could usefully do so; “not responsible” here should be taken as a para-
phrase of “responsibility is not an applicable concept in this scenario.” In other words, the question itself is at best infelicitous and at worst fallacious. For the same scenario described in psychological language, the agent could provide the group with reasons and excuses of the type that the psychological description at least suggests and for that reason a dialogue where they are discussed can be imagined to take place because of which we may have (or imagine that we would have) a reactive attitude. In reality, of course, we do not have that choice if mechanism is true, because in mechanism as in determinism it is not the agent that makes the relevant difference but facts out of his control. Either way, we cannot evaluate physical causes normatively. Finding oneself expected to evaluate physical causes is to find oneself in a contradictory situation where we have no intuitions to guide us. This also ties in with the concept of moral luck, because the intuition is that the agent who does not have libertarian free will has been lucky when his causal history has resulted in acts we count as morally good and unlucky when they are bad.\textsuperscript{16} Again, there seems to be no point in moral criticism of what is determined, or to be more precise, the criticism would carry no moral weight.

In short, determinists are not entitled to anything other than descriptive terms, and determinists who ask, answer, or just take seriously questions couched in such terms are themselves in performative contradiction. Although determinism and free will, or determinism and responsibility, may not be logically or conceptually inconsistent, the speech-act of blame is inconsistent with a belief in determinism in exactly the same way that an attempt to assert a belief in free will would be. The judge would be making a performative contradiction.

Put this way, it seems obvious that facts about the judge as well as about the actor are relevant to the actor’s responsibility. Moral questions, we are inclined to think, are questions worth asking and answering, and the judge takes himself to be doing exactly this when he stands in judgment. Of course, he may be mistaking what he is doing, so I have not shown that determinism and responsibility are incompatible; what I have shown is that, because when we take a genuine act of blame to have been committed we must take the blamer to believe that his act of blame is distinctively moral (because this was one of the internal conditions of satisfaction of the speech-act), and this is inconsistent with attributing also the belief in determinism to him, for determinism rules out the distinctively moral. Similarly and implicitly for ourselves whenever we engage in a moral practice. Now, we might try to reconcile moral
language with determinism by considering free will as a necessary and convenient fiction which is, strictly speaking, false. But even this does not get around the fact that the judgment itself, and not only the act judged, must be determined. To call something “distinctively moral” can thus be thought of as the semantic thesis that moral acts cannot be represented without the use of moral terms. If moral acts are actually performed, then certain things must be presupposed.

3 Paul and Peter

How must Paul be related to Peter and to the community?

We have already said that the standing needed for blame itself, in contrast to punishment, is fairly minimal, and in fact for blame there seems to be no real distinction between Paul and the judge. True, it is Paul that will tend to actually experience a reactive attitude, and this attitude is likely to be due more to affect than to moral reasoning of the kind I have been describing, but as I have already said the motivating reasons for the reactive attitude have little to do with the appropriateness of the blame. True, a judge in the sense of a moral and legal authority has a greater obligation to reach his decisions by reason rather than emotion than is the case of the one who is directly affected, but as long as there are reasons, that is to say, a winning strategy, the responsibility-attribution is not undermined.

There is one further thing that I think is worth saying does not undermine Paul’s reactive attitude or Paul’s thinking his attitude to be appropriate: it does not matter whether the content of the norm that Peter violates is one that Paul himself endorses. That, as a matter of fact, you have a certain entitlement to make moral demands, is a separate issue from whether you personally think that such-and-such ought to be a moral demand.

Suppose, then, that Paul rather peculiarly does not think that being stolen from ought to be a norm. Because he does not, he may not even resent Peter for stealing from him. Yet he might still, and appropriately, resent Peter for violating a norm just because it is a norm, and because as a matter of fact there is such a norm it is appropriate for Paul to hold Peter to that norm and make the moral demand of Peter not to violate the norm. Paul is no less entitled to make this demand on account of not thinking that the norm ought to be a norm. I have already said briefly in the discussion of hypocrisy that the fact that Paul might not hold himself to that norm does not undermine the responsibility-attribution either.
It is of course possible and easily understandable that in such circumstances (where he does not personally endorse or hold himself to the violated norm) Paul might withdraw the reactive attitude, but I believe that any feeling that he may have that this withdrawal is incumbent upon him on grounds of appropriateness are mistaken; such feelings are due to taking the perspective of the individual rather than the collective, where it is the appropriateness of the collective’s reactive attitude, in the form of the reactive commitment, that is relevant to the question of whether Peter should be held responsible. The moral demand gets its force from the moral community. This does not alter the fact that the moral demand is actually made by one individual to another, and that the making of such demands regulates first and foremost a particular relationship between those individuals.

As for Peter, the important thing to remember is that being a member of the moral community, and sharing goals with the moral community, in no way requires Peter to endorse those goals or to embrace those norms and expectations he is charged with violating. Nor does there seem to be any reason to suppose that Peter is unable to regulate himself according to those norms and expectations. This means that Peter is an apt object of reactive attitudes.

4 Conclusion

As I said near the start of this paper, the guiding intuition is that we take our practices of blame to have moral weight, and in doing so we find ourselves committed, whatever our other metaphysical beliefs (e.g., about whether determinism is true), to having a certain conception of moral responsibility and moral blame. I have followed a tradition that sees the blame as a speech-act and these commitments as its conditions of satisfaction.

Some of these conditions will be internal conditions, which is to say that they can be performed only by a speaker with the right psychological attitudes, and it is in this way that the speaker’s conception of moral responsibility gets embedded in the speech-act of blame. My innovation is to say that it is the moral community itself that should be regarded as the speaker, so it is the attitudes that can be attributed to the community that are relevant to the successful performance of the speech-act, rather than the individual’s attitudes. As Pettit would put it, the community must be one where reason has been collectivized, and this is a matter of following procedures.
These procedures amount to a supervenience relation between the reactive attitude attributed to the community and the beliefs of its members, or to be more precise (since we should not assume sincerity) the beliefs attributed on the basis of their speech-acts to its members (including the one to be blamed) about whether excuses or exemptions obtain. Just as this allows beliefs to be attributed to the community even in cases where no member of the community has that particular belief, so, I argued, a reactive attitude could be attributed to the community even in cases where no member of the community has that particular reactive attitude. In such cases the members of the community may be said to accept the belief or reactive attitude.

As well as beliefs and reactive attitudes, there is also an obligation to blame attributable to the community derived from the dialogical interplay of reasons as a corollary to the belief that the agent is responsible. In this case there will be a similar obligation that is attributable to its members, or at least those in the community with moral authority. Even though they have the same content, the obligations are nonetheless distinct because some considerations can defeat or override some or even all of the obligations of the individuals without defeating the obligation of the community. Even so the community’s obligation supervenes on the same set of propositional contents; the belief that the agent is responsible (i.e., has no winning strategy) comes first and the obligation to blame is its corollary (provided that the agent is responsible to the community and consequently obliged to provide it with a justification in the first place).

It follows because of this that the community’s members have obligations, so it is still true to say that the moral obligations of the community supervenes on those of its members, even if the order in which we actually attribute obligations tends to go from the community first to the members second. Even though all the member’s obligations can be defeated, a defeated pro tanto obligation is still a pro tanto obligation, so there will always be pro tanto obligations in the supervenience base and it is not implied that the community must be strongly rather than weakly autonomous. This is how it goes with acceptances too, since, provided that certain other conditions on membership in the community are satisfied, there is a pro tanto obligation for those in the community to accept the belief attributed to the community and others in the community are entitled to expect certain performances on the basis of that obligation, such that they are entitled to require you to justify yourself if either you do not carry out such performances, or (as is more likely the case) you
do something that conflicts with that obligation. When the acceptance is that some agent is responsible and there is a corresponding obligation to blame him, and I do not, this does not mean that the agent is not to blame but simply that I am accountable for not blaming him. There is then two cases of moral expectations not being met, by the agent and then by myself.

Because there is this distinction between beliefs and acceptances, and moral obligations with respect to both, there may be a conflict between what you personally believe to be the morally right thing to do and what you accept as the morally right thing to do, and there will be a corresponding conflict between obligations. On the basis that you have violated an expectation you can be required to give a justifying reason, and you may say that your moral conscience did not allow you to, that for you the pro tanto obligation was defeated and consequently not your all things considered obligation. Often we admire such people, especially when hindsight has shown them to be right (our admiration is probably biased by a one-sided diet of examples, although it commendably cautions us not to take our moral superiority for granted). Whether and to what extent this is acceptable as a justifying reason to the moral community to whom it is given, however, is a matter of the substantive ethical code of that community. Some form of mitigation would probably be appropriate, but this need not undermine the blame or responsibility-attribution itself. If, in the exchange of reasons, the agent fails to give a reason considered by the moral community to be justifying, then on my account he is (considered to be) morally responsible. One can, then, be morally blamed for doing something that is morally good, and one can be morally obliged to do something that one considers to be morally bad. The obligation to blame is pro tanto only, but it remains: this is captured in my definition of accountability.

In summary, the community must be (weakly) autonomous: it must be possible to attribute intentional and moral properties to the community that supervene on those of the community’s members but can vary in propositional content and/or being defeated or not (which affects in consequence what comes out as all things considered).

Some of the conditions of the speech-act of blame will be external conditions. The conditions on membership in the community, whose function in moral practices and moral justification are described above, are such external conditions. I have used the accounts of Tuomela and Kutz to cash out these conditions. The details do not actually matter too much and one can probably substitute other accounts without loss.
The only thing that we need to ensure is: i) that a distinction between an acceptance and a belief is maintained (i.e., reason is collectivized); ii) that beliefs attributed to the group are normatively binding on the group; iii) that at least sometimes this normativity is moral, and; iv) that corresponding to this moral normativity are pro tanto moral obligations. To these we may add: v) that the procedure from which the acceptance results embodies norms of communicative rationality, or to put it in dialogical terms, that the collective follows rules of dialogue that allow for the “free” exchange of reasons, the commitments in the dialogue corresponding to the attitudes we attribute to the group, of which group both the accuser and the accused are parts and play complementary roles in the dialogue, even though on another level they are mutually antagonistic.19 The accountable agent must be given the opportunity to answer, and reason cannot be collectivized in such a way that excludes certain individuals or sub-groups within the collective.

Moral reasoning – or at least reasoning about whether an agent is morally responsible – is reasoning about who has the winning strategy in a dialogue conducted according to rules that allow for an appeal to dialogical commitments shared by the moral community, and where the substantive content of these commitments can be fitted into the Strawsonian framework of excuses and exemptions. In my view, this is a framework only, and different moral communities may have different substantive views on what fits into this framework, that is to say, on what they are prepared to consider as grounds for the excuses and exemptions. One reasonably substantive thing that I think must count as grounds for an exemption is moral bad luck. I have taken the view that this is embedded in the community’s conception of responsibility itself, but perhaps it might just be an empirical fact that all communities accept an ‘ought implies can’ principle, and from the way I have justified it – that morally blaming the morally unlucky would be pointless – it does not follow that it is a conceptual truth about responsibility, or that we must consider it to be such when we blame.20 There are many speech-acts that we are able to perform, and even do perform, even though performing them is pointless or redundant. Pointlessness does not equate to a performative contradiction, as perhaps it would do when we blame someone who we know to be quite unable to understand what it is we are doing or to recognize our illocutionary intentions, as to someone for whom an exemption obtains ([26]).

Blaming is a collective act, and its appropriateness depends on the collective’s substantive ethical concerns. Responsibility is definable in
terms of performance of this act. If one of the external conditions is not satisfied then the blamer does not have the standing to blame; the blame is not successfully performed and is not binding on the one who is blamed, just as a legal sentence must be pronounced by a judge or other authorized party or else it is not a legal sentence at all. The external conditions are not difficult to satisfy, however. Also, although Paul, in the example, is the one who is likely to have the reactive attitude, and may express the reactive attitude, he is not in a privileged position with regard to blaming Peter, and in fact it is only because Paul takes his blame as having moral weight, that is to say, by integrating into his experience the perspective of the community itself that is the source of the moral norms Peter has violated, that Paul’s blame has the particular kind of moral force that Paul takes it to have. Again, this integration of the two perspectives is not a difficult condition to satisfy, but is in fact a commonplace of moral experience, and both Paul’s own reactive attitude and that reactive attitude we attribute to the community are expressed in one and the same act, though it is the latter that accounts for the moral force.

Finally, I claimed that those who blame see their blame as the appropriate moral response to a moral wrongdoing, and thus as requiring a moral justification by moral reasons; Strawson agrees with the incompatibilists that those compatibilists change the subject who give an instrumental justification in terms of management, regulation, and put objective attitudes in place of the reactive attitudes, thereby relegating reactive attitudes to mere psychological phenomena that provide no insight into moral questions as such. This is just not how we conceive of our moral practice, and our moral conceptions must reflect this fact; these conceptions underlie our blaming and are implicitly affirmed when we blame. As already said, we incur a commitment to these conceptions, and even if we cannot formulate them, the patterns of our reactive attitudes seem to show that we are at least sensitive to the features involved in them. I put this in terms of the “distinctively moral.” I argued then that if there was a sphere of the distinctively moral – and our moral practices seemed to commit us to there being one – then determinism must be false, for determinism cannot tolerate any genuinely normative language but must re-interpret it as a disguised description of our psychology. Of course, on one level it does describe our psychology, but we do not mean or intend it that way when we use it. When we use normative language, and especially moral language, either directly or through practices indirectly analysable as speech-acts of praise or blame, we are also committed
to determinism being false and are in a performative contradiction if we blame while believing determinism to be true, irrespective of whether we or anyone else are aware of this. This was not so much because an agent’s actions being caused amounted to coercion or some other reason that might undermine his moral responsibility, but because we cannot conceive of the act of blaming itself as being caused and yet still having the moral force and weight that we take it to have. Of course, as a matter of fact the act of blaming might be caused without our being in any way aware of it, so determinism itself is not inconsistent with holding people responsible. What I do say is that the conception of responsibility that we find ourselves committed to excludes the possibility of determinism being true, so it is a pragmatic/performative contradiction to allow that it is true or even that it might be true when we blame.

Acknowledgements

This work was supported by two grants of the Portuguese Foundation for Science and Technology (FCT): “Is moral reasoning essentially dialogical?” (SFRH/BPD/77687/2011) and “Argumentation, Communication and Context” (PTDC/FIL-FIL/110117/2009).

Notes

1 It is this intuition (viz., that this kind of compatibilist has taken the morality out of our presumptively moral practices of blame and punishment) about which Strawson [22, 4] says that the ‘optimist’ (Strawson’s term for the Schlickian kind of compatibilist) has nothing more to say, to compatibilism’s detriment. It should be observed that Strawson limits the discussion of compatibilists to the optimists, apparently ignoring the other kinds of compatibilism just mentioned.

2 I must admit that I always feel reservations about the interpretation of Strawson’s arguments as drawing conclusions about the concept of responsibility itself. I prefer to think of reactive attitudes theory as an account, or perhaps a methodology for acquiring an account, of responsibility-attribution, and Strawson’s arguments as making epistemological points rather than conceptual points. I would also question whether Strawson really says that being responsible is a matter of being held responsible; all that I think strictly follows from his arguments (if they are good) is that the reasons for which we consider someone to be responsible can be discovered by naturalistic description of the pattern of our reactive attitudes and do not require further justification.

3 Excuses function by showing that an action did not manifest ill-will and obtain when the agent could say in response “I didn’t mean to,” “I hadn’t realized,” “I didn’t know,” along with things like “I couldn’t help it” when backed by “I was pushed,” “I had to do it,” “It was the only way,” or “They gave me no choice.”
One kind of reason that the norms of appropriateness are sensitive to, then, are those concerned with quality of will. Importantly, an agent excused in this way is still held to expectations and as one for whom reactive attitudes are apt [22, 7-8].

Exemptions function by showing that the agent’s agency is too diminished either by their incapacity to understand expectations and norms, or by their incapacity to regulate themselves by those expectations and norms, for it to be appropriate to hold such agents to those expectations and norms; they are not apt objects for reactive attitudes. Exemptions obtain when the agent (or someone on the agent’s behalf) could say in response “He wasn’t himself,” “He has been under very great strain recently,” “He was acting under post-hypnotic suggestion,” “He’s only a child,” “He’s a hopeless schizophrenic,” “His mind has been systematically perverted,” or “That’s purely compulsive behaviour on his part.” A second kind of reason that the norms of appropriateness are sensitive to, then, are those concerned with incapacity [22, 8-10]. Here the reactive attitudes should be replaced with objective attitudes and by doing so we treat the agent as less than fully human, in some sense, and seek only to manage him.

When an excuse or exemption obtains, we do not consider resentment appropriate. Indignation is the analogue of resentment when one is not oneself the injured party, and is argued (on the grounds that supposing otherwise leads to a pathological kind of moral solipsism in which one considers oneself to have some kind of morally privileged position compared to everyone else) to have the same conditions on appropriateness as resentment.

As an example, consider a tenure committee from the point of view of the members’ beliefs about the areas in which the candidate excels. 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-$q$ 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 these conjuncts that they are 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. Proposition-wise supervenience, where the group’s belief about some proposition (in the example, $g$) is taken directly as a function (in this example a majority, but other functions are possible but lead to the same results) of the members’ beliefs about that same proposition, leads to refusal of tenure and irrationality at the group-level [12, 92-93].

[9, 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).

6 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.

7 There is a strong similarity between my dialogical view and Wallace’s [25] view that an actual reactive attitude is not necessary to hold an agent responsible if instead there is a judgment that a reactive attitude would be appropriate.

8 [20, 470] argues:

From the fact that I do not “hold” someone responsible for an objectionable attitude or action, in the sense of actively blaming her for it, it might be inferred that I do not hold her to be responsible for that thing, in the sense of open to moral appraisal, or that I do not hold her to be culpable for it, in the sense of open to legitimate moral criticism. But these things simply do not follow. If a good friend of mine is under a lot of stress, for example, I may not “hold” her responsible, in the sense of actively blaming her, for some insensitive comments she makes to me. I can judge both that she is responsible for her comments, and that she is open to legitimate moral criticism for them (because they are hurtful). But given the circumstances, I may decide that it would be uncharitable for me to take up attitudes of anger and resentment, or to explicitly reproach her in any way. In making such a judgment, however, and in renouncing these attitudes and responses, I need not think that my friend is not really responsible, or not really at fault, for her behavior.

9 [11] reports empirical research that shows how widespread this intuition is, although she considers it to be a tu quoque. A female student of Libby’s argued thusly:

If I think that it is likely that I would have owned slaves had I been born into the American slave-owning South, then it is not appropriate for me to say that slave-owners were blameworthy. Likewise, if I think that I would have followed the orders of the SS, then it is not appropriate for me to blame those who did. [11, 10]

However, Libby considers this intuition mistaken:

Perhaps she was worried that she would be labeled a hypocrite for pointing out the moral mistakes of others in light of the fact that she herself had surely committed her own mistakes. . . . [T]his charge of hypocrisy . . . would be fallacious. The fact that Emmy has failed to perform all and only morally permissible actions is not logically relevant to her arguments attempting to establish that some other individual has done something morally impermissible. [11, 7-8]
10 According to [8] it is not necessarily inappropriate to feel guilt or shame for practices that were in fact socially approved of at the time.

11 This over-simplifies. For a more subtle analysis see [1].

12 McKenna calls such a dialogue a *moral responsibility exchange*. In [4] I show that this follows closely a particular dialectical model regulated by a specific set of rules. Note also that another way of stating the same condition is to say that G has the reactive attitude for resentment towards A, for the condition here is the same as that by which resentment is attributed to the collective.

13 We can take the same approach to the personal *pro tanto* obligations mentioned earlier. Does the obligation of caring for a sick child defeat the obligation to show due diligence? I think it depends on the substantive ethical norms of the moral community whether this is to count as grounds of, for instance, the excuse “I had no choice.”

14 “The factual . . . is a paradigm case of the non-normative,” Cohen [5, 316] says. This does not rule out that our methodology for finding the answer may be based on description. It does not rule out a Strawsonian kind of naturalism where the conditions of responsibility are held to be discoverable by investigation of and description of our psychological habits regarding reactive attitudes.

15 Nahmias et al have done the experiment of describing the same deterministic background and varying whether an action is described as caused by his intentional states or by his neurophysical states, notwithstanding the fact that each are equally determined. They discovered that respondents were ready to attribute responsibility where psychological language was used but not when physical language was used. This seems to show that people are generally amenable to compatibilism between determinism and responsibility, contrary to the claims of incompatibilists, but not to the compatibilism between mechanism (the reduction of the description to a physicalist language) and responsibility. When examined more closely the folk conception of responsibility is not incompatibilist, it is argued, but only seems to be [16]. Against this, [18] argued that the folk conception of responsibility was incompatibilist and that what the studies (backed up with studies of their own) showed was that when presented ‘concrete’ cases subjects tended to exhibit compatibilist tendencies, but that when presented abstract cases subjects tended to exhibit incompatibilist tendencies. This in turn was explained by the fact that concrete cases elicited high affect which biased the psychological process; compatibilist intuitions were performance errors. But [17, 255-256] turns this around:

With no other information given, people tend to assimilate determinism to coercion, but this suggestion is concealed when an intentional action is described in concrete terms. Determinism is also sometimes assimilated in people’s minds to reductionism. But determinism does not preclude intentional, rational action. The concrete case avoids these faulty assimilations.

These assimilations being faulty in her opinion, it is incompatibilist intuitions that are performance errors.

16 See [3, 26-27] where I argue that the issue of moral luck underlies and explains many of our otherwise puzzling intuitions.

17 [7, 9] will similarly claim that moral demands are second-personal though they get their force from being norms of the moral community:
I claim that to understand moral obligations as related to moral responsibility in the way we normally do, we have to see it as involving demands that are “in force” from the moral point of view, that is, from the (first-person plural) perspective of the moral community. . . . [T]his does not diminish their second-personal character, since this concerns their “demand addressing” quality.

18 A situation that is similar in some ways occurs when your conflicting belief is due to enculturation in a different moral community. Also, because one may belong to more than one moral community, acceptances may conflict with other acceptances as well as with personal beliefs; the reasons that would be acceptable in one community need not be acceptable in another. What would be the justifying reasons in such cases, though, are not those reasons themselves, but the fact that one has been enculturated to accept and comply with those reasons. This is a separate reason, and whether it is acceptable is a substantive issue and cannot be settled by philosophical analysis ([1]).

19 There is a hint of circularity here. Is there an obligation to allow the accountable agent the opportunity to defend himself? If there is, as I suggested earlier, then it does not seem that this obligation can be brought about as the result of a dialogue of this kind, for this presupposes such an obligation. Here, then, I say only that the collective actually follows the rules, not that they are obliged to. One wonders whether this is strong enough, for then it is not obvious that one can be sanctioned for not following the rules of dialogue and the blame will consequently be deprived of its moral force. My preference is to see the rules of dialogue as having a moral force of their own, and this because dialogue is itself a co-operative activity in which each participant is entitled to expect certain performances from every other, and a participant cannot unilaterally break off this activity once there is de facto agreement to engage in it. My concern with this, though, is that this seems to give moral force to the rules just because they are rules, because they have been agreed to or because they are instrumental to the goal of the activity, and not because of their content as such — it seems quite possible that completely different rules might have been agreed to, and this account would give them the same moral force. What I think I would need to show is that such norms are instrumental, not to the specific goal, but to co-operation as such, or perhaps to co-operation whose goal is to resolve an issue that is “distinctively moral.”

20 Perhaps we might offer a Watson-type argument that reactive attitudes would not have a communicative function for the morally unlucky. Watson insists that impossibility of having a communicative function is not just an instrumental justification, but I find him to be ambivalent on this point. In cases where the agent is incapable of recognizing moral reasons I think Watson is right and the justification is not instrumental, but in the case where the agent acts out of character due to stress, the inappropriateness of the reactive attitude does seem to be justified on the instrumental grounds that doing so would be pointless ([26]). The morally unlucky person does not generally fail to recognize moral reasons, and so seems more like an example of the second case than the first.
References


Austin on Literal Meaning

ODAI AL ZOUBI

Abstract

In this paper, I examine the debate between Alice Crary and Nat Hansen concerning Austin’s view of ‘literal meaning’. Crary suggests that Austin thinks that there is no literal meaning, while Hansen thinks that for Austin there is literal meaning. I will argue that for Austin there is indeed a literal meaning, a fixed meaning, which sentences carry across all occasions of use, however, such meaning does not suffice to determine whether, independent of these occasions, the sentences can be used to say something true or false. Thus I suggest a middle ground between the two readings. My reading highlights the importance of Austin’s claim that ‘true’ is a dimension-word, which both Crary and Hansen overlook.

Keywords: J. L. Austin, truth, literal meaning, speech acts

In this paper, I analyse a recent debate between Alice Crary and Nat Hansen concerning Austin’s view of ‘literal meaning’. The question which Crary and Hansen debate is whether Austin thinks that there is a literal meaning of sentences, namely what is true or false and is invariant over different uses. Crary thinks that he doesn’t and Hansen thinks that he does. I will seek to show that the answer lies somewhere between these two poles: I will argue that, according to Austin, (at least some) non-ambiguous sentences have a fixed meaning, which they carry across all occasions of use. Such sentences meaning what they do, however, does not suffice to determine whether, independent of these occasions, the sentences can be used to say something true or false. This interpretation stands in contrast to Crary’s interpretation in so far as Crary denies that sentences have such a fixed meaning; it stands in contrast to Hansen’s interpretation in so far as Hansen affirms that sentences having such fixed meaning suffices to determine whether, independent

http://www.kriterion-journal-of-philosophy.org © 2016 The author
of occasions of use, the sentences can be used to say something true or false.

My key claim is that both sides of the debate ignore one of Austin’s most important findings: that ‘true’ is a dimension-word. Whilst both parties represent some aspects of Austin’s position correctly, their failure to ramify fully his view that ‘true’ is a dimension-word undermines their conclusions. I will show that the identification of ‘true’ as a dimension-word is central to Austin’s theory of speech acts and fundamental to his view on literal meaning.

Firstly, in section 1, I will introduce Crary’s and Hansen’s arguments. Next, in section 2, I will explain what a dimension-word is and how Austin takes ‘true’ to be such a word. I will then argue that Austin’s view of ‘true’ as a dimension-word plays a central role in introducing his theory of speech acts. In particular, as we shall see in 3, it influenced the development of that theory, ultimately forcing him to abandon his distinction between performatives and constatives. It is only against this background that we will be able to understand the relation of truth to the two parts of the speech act, the locutionary and the illocutionary, as we shall see in 4, and then we will go back to the Hansen-Crary debate on Austin’s views of ‘literal meaning’. Finally, in 5, I will summarise my conclusions.

1 Crary and Hansen on Austin on literal meaning

Crary gives a reading of Austin which portrays him as attacking the view of ‘literal meaning’. According to Crary, Austin tries to show that the ‘traditional statement’ is an illusion.

The picture of correspondence between language and the facts that Austin takes to be implicit in a traditional ideal of the ‘statement’ is one on which the business of corresponding to the facts is the prerogative of what might be called bi-polar ‘statements’ or propositions, i.e. ‘statements’ or propositions that always describe states of affairs either truly or falsely.1

Crary continues ‘[Austin] proceeds by arguing that this idea [the traditional statement] is nourished by a view of meaning on which sentences possess what are sometimes called literal meanings (i.e. meanings they carry with them into every context of use) and by arguing that this view fails to withstand critical scrutiny.’2
Crary’s argument can be divided into two steps. In the first, Crary maintains that

Austin stresses that he thinks that whenever I say anything . . . I perform both a locutionary act ... and an illocutionary act... He is drawn towards this view by the thought that there is no such thing as identifying the meaning of a combination of words (or: no such thing as identifying the ‘locutionary act’ performed when a combination of words is uttered) independently of an appreciation of how those words are being used to say something to someone on a particular occasion (or: independently of an appreciation of their ‘illocutionary force’).3

In the second, Crary takes the interrelated connection between locutionary and illocutionary acts to show that there cannot be a literal meaning of sentences.

Austin’s account of locutionary and illocutionary acts, taken as a whole, brings into question the idea that we might develop a theory that could be used to identify the locutionary acts performed whenever particular sentences are used... he criticizes it by rejecting as flawed an idea that it presupposes, viz., that it is possible to isolate the locutionary act that is performed when a particular sentence is employed in the absence of a grasp of the illocutionary force with which it is being used.4

According to Crary, then, Austin thinks that the sentence cannot carry with it an invariant literal meaning in all its uses because, in order to understand an utterance, the whole speech act has to be grasped. Because we can’t separate the two acts, because we need to understand the speech act as a whole, it is not possible to understand the ‘locutionary meaning’ in isolation from the whole speech act, and there therefore cannot be an invariant literal meaning. So Crary doesn’t just deny that sentences, independent of occasions of use, can be used to say something true or false; she also denies that such sentences have fixed meaning, independent of these occasions.

Hansen disagrees with Crary. He argues that Austin seems to endorse the literal meaning view, where literal meaning ‘is that grasp of the meanings of words and the rules by which they are combined into complex expressions (including sentences) [which] enables one to know what has to be the case in order for the sentence to be true.’5
Firstly, Hansen maintains that Austin thought that locutionary acts could be separated from illocutionary acts. ‘Austin nowhere explicitly commits himself to the idea that identifying the locutionary act performed by an utterance requires an appreciation of the illocutionary act performed by that utterance as well.’ He quotes Austin saying that

[it] might be perfectly possible, with regard to an utterance, say 'It is going to charge', to make entirely plain 'what we were saying' in issuing the utterance... and yet not at all to have cleared up whether or not in issuing the utterance I was performing the [illocutionary] act of warning or not. It may be perfectly clear what I mean by 'It is going to charge' or 'Shut the door', but not clear whether it is meant as a statement or warning, etc.

If Hansen is correct, and we can understand the locutionary meaning in isolation from the illocutionary force, it might be possible to read Austin as allowing that a literal meaning might be assigned to the locutionary part.

But Hansen, secondly, goes further. According to him, even if you have to understand the whole speech act in order to understand the locutionary meaning, this does not entail that there is no ‘literal meaning’. As long as there is a distinction between ‘locutionary meaning’ and ‘illocutionary force’, and as long as we think that there is a way to make this distinction clear, it is possible that different speech acts might have different ‘forces’ and still share the same ‘meaning’. He maintains that there is nothing in Crary’s argument which blocks this approach and that, as a result, it is plausible that this shared ‘meaning’ is indeed invariant across different uses, and is what is true or false.

Hansen doesn’t just affirm that sentences, independent of occasions of use, might have fixed meaning; he also affirms that this fixed meaning suffices to determine whether the sentences can be used to say something true or false, independent of these occasions.

However, both Crary’s and Hansen’s accounts of what is ‘true’ or ‘false’ and the locutionary/illocutionary distinction is problematic. I will argue that Austin’s view lies in between these poles: for Austin, sentences have a fixed meaning, which they carry across all occasions of use, however, such fixed meaning does not suffice to determine whether, independent of these occasions, the sentences can be used to say something true or false.

The interpretation I propose is compatible, in its main lines, with Charles Travis’s influential views of truth and meaning. Travis dis-
tungishes between two kinds of what he calls semantic properties for sentences. ‘The first sort of property is one of relating in a given way to truth (or falsity). Properties of being true (false) if, given, of, or only if, thus and so, or thus, or the way things are, are all within this class... The second sort are properties identified without mention of truth, and on which truth-involving properties depend’t. According to Travis, there are some properties, what we can describe as a level of meaning the words (sentences) would have independent of being true or false. However, there is another level of meaning, other properties, which relate to judgements of true or false. Travis then distinguishes between the fixed meaning sentences carry across all occasions of use, on the one hand, and what is to be true or false on the other. The latter is to be identified according to a specific speaking of the words, in occasions of use. I find Travis’s view closer to Austin than Crary and Hansen, and I will explain why in this paper.

Travis does not intend to give an explicit interpretation of Austin’s text nor does he cash out his view in terms of locutionary meaning and illocutionary force, as I will do, he nevertheless acknowledges Austin as an inspiration. In the introduction to his Occasion Sensitivity, he quotes Austin on the relation between sentences and truth,

‘It seems to be fairly generally realized nowadays that if you just take a bunch of sentences... impeccably formulated in some language or other, there can be no question of sorting them out into those that are true and those that are false; for (leaving out of account so-called ‘analytic’ sentences) the question of truth and falsehood does not turn only on what a sentence is, nor yet on what it means, but on, speaking very broadly, the circumstances in which it is uttered. Sentences are not as such either true or false.’

The difference between my interpretation and Travis’s view lies in my attempt to give an explicit interpretation of Austin’s text, and in my claim that Austin’s view that ‘true’ is a dimension-word is the key to understand Austin’s views on truth, a claim Travis doesn’t give the crucial role I give.

To understand Austin’s position on literal meaning, I start by examining his claim that ‘true’ is a dimension-word. To avoid any misunderstanding, I want to make it clear that in the following I distinguish between sentence-meaning, which is the subject of the debate on literal meaning as explained above, and word-meaning. My discussion in section 2 is restricted to word-meaning, in this case the word ‘true’.
argument is that Austin’s claim that true is a dimension-word, i.e. that it has different but related meanings, as will be explained below, is crucial to understand his views of sentence-meaning. However, we need to keep the two accounts, the account of Austin’s view of word-meaning and his view of sentence-meaning, distinct all through. This will be clearer below.

2 True as a dimension-word

Austin distinguishes between two kinds of words: words that have one meaning, and words that have multiple, unrelated meanings. In his examination of the different uses of ‘real’, Austin points out that this word ‘does not have one single, specifiable, always-the-same meaning... Nor does it have a large number of different meanings – it is not ambiguous, even “systematically.”' According to Austin, there are words that have always-the-same-meaning, like ‘yellow’ or ‘horse’, and, on the other hand, there are ambiguous words like ‘bank’, which can mean either a financial institution or the edge of a river. These are completely different meanings. There is, nevertheless, a middle ground between these two kinds of words. He writes; ‘If we rush up with a demand for a definition in the simple manner of Plato or many other different philosophers, if we use the rigid dichotomy “same meaning, different meanings”... we shall simply make hashes of things.’ Many philosophers neglect this middle ground, he thinks, and, as a result, they fall into a false dichotomy: ‘one meaning/ambiguity’, which often causes them erroneously to look for one meaning for each word.

In particular, Austin argues that with certain types of word that have multiple meanings there might be something in common between all the uses of the word, but that this commonality exists at an ‘abstract’ level, and that focusing on this common factor obscures the many differences that exist at the ‘concrete’, contextual level. In other words, the meaning of these words involves two levels: what we might term ‘abstract meaning’/’semantic function’ and ‘specific meaning’. The former, in virtue of being abstract, might well be consistent across uses of the word in different contexts and cases, whereas the latter is likely to vary depending on the circumstances and contexts in which the word is used. To use today’s jargon, we might say that these two levels of meaning correspond to the semantics-pragmatics division: abstract meaning is semantic, and specific meanings are pragmatic.

We can say that according to Austin, we have three types of words:
unambiguous words, which always have the same meaning, like ‘yellow’ and ‘horse’, and ambiguous words, which have two or more totally different and unrelated meanings, like ‘bank’. The third type of word is words that have different but related meanings: multiple-related-meanings words like ‘true’, ‘real’ and ‘freedom’. There are many different kinds of words of this type; one prominent type is ‘dimension-words’. According to Austin, ‘true’ is a dimension-word: as we will explain below, its abstract meaning is to appraise the relation between words and the world, its specific meaning is to specify this appraisal in different ways: ‘accurate’, ‘not very good’, ‘too concise’, ‘general’... etc.

Things would be clearer after we examine some examples below. The dimension-word ‘is the most general and comprehensive term in a whole group of terms of the same kind, terms that fulfil the same [semantic] function.’ For example, according to Austin, ‘real’ is the most abstract word in a group of words which have the same semantic function, the same abstract meaning. Members of this group of words, ‘on the affirmative side, are, for example, “proper”, “genuine”, “live”, “true”, “authentic”, [and] “natural”; and on the negative side, “artificial”, “fake”, “false”, “bogus”, “makeshift”, “dummy”, “synthetic”, [and] “toy”.’ (Austin, 1962, p. 71). Thus, ‘real’ picks out a set of words which all possess the same semantic function, the abstract meaning ‘real’, but which are also individually chosen in their different forms in order to convey the particular sense of ‘reality’ (or its opposite) appropriate to the specific context in which they are used.

Dimension-words define a semantic dimension and the range of terms appropriate to the particular abstract meaning or semantic function of the particular dimension-word. The dimension-word could, in fact, substitute for any of the members of the family of words within its dimension in virtue of all members possessing this abstract meaning along with their own context specific concrete meaning. However, the necessarily abstract nature of the meaning of the dimension-word means that its usage in particular situations would be unlikely to convey the required specificity of concrete meaning. Thus, although the abstract meaning/semantic function of all of the terms in one family is the same and is constant in all the uses of a dimension word, Austin wants to show, that identifying this common thing and focusing on it will not provide a sufficiently robust or accurate basis on which to determine meaning. The semantic function is too thin; it needs to be supplemented by the specific meaning, which is to be changed according to the context.

It is the combination of the shared abstract meaning and the context-
related specific meaning which means that dimension-words don’t have one meaning in all of their uses, and yet are not ambiguous. Rather, they have a number of different-but-related specific meanings which are unified by their common possession of the ‘abstract meaning’ of the term.

Austin thinks that ‘true’ is a ‘dimension-word’, in virtue of which it has something in common in all of its uses, what we called the ‘abstract meaning’/semantic function, but no one specific meaning in all of its contexts or circumstances of use. The semantic function associated with ‘true’ fulfils the following purpose: ‘True’ and ‘false’ are just general labels for a whole dimension of different appraisals which have something or other to do with the relation between what we say and the facts. In addition, he notes that the different terms which belong to the family, and share this semantic function, are quite diverse. Thus, we find within its ambit terms such as ‘exaggerated’, ‘vague’, ‘bald’, ‘rough’, ‘misleading’, ‘not very good’, ‘general’, ‘too concise’, ‘fair’... etc. These are the terms which we, in ordinary language, use for the appraisals of utterances. All members of the family share the same semantic function, to assess the relation between utterances and the world, but differ from each other in other aspects and characteristics.

According to Austin, it is rare that we use ‘true’ or ‘false’ in ordinary language. Austin, as we shall see, thinks that ordinary users employ these abstract terms only in logic and mathematics. Instead, we tend to pick a member of the family (such as ‘exaggerated’ or ‘vague’) that better represents the particular aspect of truth or falsity appropriate to the situation. Philosophers, however, are inclined to do the opposite and focus only on the two most abstract terms in their discussions, and ignore the other terms of the family more suited to normal or ordinary cases.

To illustrate the point, let us examine an example from Austin: ‘the galaxy is the shape of a fried egg’. He encourages us to ask whether this statement is true or false. His view is that we cannot tell if this statement is ‘true’ or ‘false’ independent of the purpose of making it and its audience. Thus, if the context and audience are such that a rough or approximate similarity will suffice, such as might occur in a discussion with young children about shapes of different star systems, then we may agree that the statement is ‘true’. Similarly, if we were talking to an astrophysicist who would base a number of precise calculations on our answer, we will almost certainly have to reply that the statement is ‘false’.

This example thus supports his underlying claim that ‘true’ is a
dimension-word. As we have seen above, we are generally able to answer the question concerning the truth of the claim only if we are given sufficient context. In addition, it seems equally clear that, in ordinary language, we would indicate any contextually determined lack of precision by using a more specific word from within Austin’s putative family of words in the ‘true’ dimension. Thus, in the circumstances of discussing the shapes of star systems with small children, we might well say that the galaxy is ‘like’, or ‘roughly’ the shape of a fried egg. This seems to indicate that the range of words Austin’s ‘true’ dimension can be used, from the abstract through to the concrete, depend on the purpose, audience and context of the utterance.

Let us take another example from his explanation on ‘real’. The difference between the dimension word ‘real’ and the less abstract terms in its family of words is the following: ‘the less general terms on the affirmative side have the merit, in many cases, of suggesting more or less definitely what it is that is being excluded; they tend to pair off, that is, with particular terms on the negative side and thus, so to speak, to narrow the range of possibilities.’ Austin offers some examples of this; ‘[I]f I say that I wish the university had a proper theatre, this suggests that it has at present a makeshift theatre; pictures are genuine as opposed to fake, silk is natural as opposed to artificial, ammunition is live as opposed to dummy, and so on.’ In each of these cases, that which is excluded is more clearly defined, and the intended meaning better captured, because the more specific and more concrete member of the family is used rather than the more abstract dimension-word.

The abstract word can substitute the more specific word in any context. Note that the substitution works only in one direction: For example, in assessing an utterance as ‘fair’, what we are doing can always be done by using the more abstract word, the dimension-word, ‘true’; however, if we want to assess an utterance as ‘true’, we cannot use ‘fair’ instead of ‘true’ in any context, but only in some specific contexts. In addition, you cannot substitute one of the specific words of the family with another specific word.

The context-independent abstract level (in this case ‘true’) is generally too thin to carry the burden of expressing specific semantic meaning, hence its substitution in specific circumstances by the sorts of words Austin lists. However, the dimension-word, in virtue of being the most abstract representation, can in principle substitute for any of the specific words in its family, but it is clear from these examples that its effectiveness in conveying meaning in such circumstances will be limited.
The problem, as Austin sees it, is that philosophers focus on the abstract words and ignore the more specific ones. In other words, they always substitute the abstract words for the specific ones. Austin looks at the words we use in everyday life and finds that we generally use the more specific ones while the more abstract words are only used rarely. He, thus, gives us a new insight into the working of dimension-words, and into their different, but related, meanings. This insight is crucial to understand his theory of speech acts, and his views of literal meaning and sentence-meaning, which we will discuss in the next sections, starting with the performative/constative distinction.

3 The performative/constative distinction and its collapse

In his earlier writings, such as ‘Other Minds’ and ‘Truth’, Austin proposes that we can distinguish between ‘performatives’ and ‘statements’. I will call this the ‘constative/performative doctrine’. In How To Do Things with Words, ‘Performative Utterances’ and ‘Performatives-Constatives’, however, Austin later finds that the distinction is instable, expressly because ‘truth’ is a dimension-word, and he comes to realize that a new theory of speech acts is needed as a result. In what follows, I trace his thought through this development, starting with the performative/constative distinction and its collapse.

3.1 The performative/constative distinction

According to Austin, philosophers used to take every utterance of the declarative grammatical form (an utterance which is a not question, command... etc.) to describe states of affairs, or report or state facts. As a result, they thought that they must be either true or false. Other utterances, which don’t take the declarative form, such as questions or commands, are not ‘true’ or ‘false’. Let us call utterances which are either true or false ‘statements’. However, says Austin, ‘it has come to be realized that many utterances which have been taken to be statements (merely because they are not, on grounds of grammatical form, to be classed as commands, questions, etc.) are not in fact descriptive, nor susceptible of being true or false.’ Austin observes that an utterance which takes the declarative form is not a statement ‘when it is a formula in a calculus: when it is a performatory utterance: when it is a value-judgement: when it is a definition: when it is part of a work of fiction’. These are different kinds of utterances: they take the declarative form,
but are not descriptive. One such important kind of utterance is the ‘performative’.

According to Austin, the distinction between performatives and constatives is as follows.\textsuperscript{29} Constatives are utterances which are either true or false. For example, when you state something, or describe something, or report something, your utterance is either true or false. Take for example ‘the cat is on the mat’. This is a declarative sentence, which is descriptive. It describes how things are, and it is true or false, if the states of affairs are, or are not, as it states.

In uttering a performative, on the other hand, I do not describe a state of affairs, or report something, and my utterance cannot be taken to be true or false. Instead, I do something. For example, in a marriage ceremony, when I say ‘I do’, ‘I am not reporting on a marriage: I am indulging in it’;\textsuperscript{30} or when in some official ceremony I am supposed to name a ship, I say, ‘I name this ship the Queen Elizabeth’; or when I say ‘I bet you sixpence it will rain tomorrow’. Other examples include: ‘I promise that ...’ and ‘I apologize’. Thus, in uttering a performative we get married, or name something, or promise, or apologize. What we say is not true or false, and we don’t state, or describe, or report anything. We do something else.

However, simply uttering a performative is not sufficient to constitute the specific act. Saying a few words is not marrying: ‘The words have to be said in the appropriate circumstances.’\textsuperscript{31} One way to highlight this dependence on appropriate circumstances is to consider how we might fail in doing the act. For example, if I am married already, then saying ‘I do’ in the ceremony, will not make me married. If I am not the person who was chosen to name the ship, then saying ‘I name this ship...’ fails: the ship was not named, even though I uttered the words; and if no one wants to bet with me, then I haven’t bet anyone. In each of these situations something goes wrong because some factor in the context is inappropriate. In such circumstances, according to Austin, the act is ‘to some extent a failure: the utterance is then, we may say, not indeed false but in general unhappy’.\textsuperscript{32}

However, in the next section we will see that Austin comes to realise that the constative/performative distinction is instable, and in accordance with its collapse, he offers his theory of speech acts.
3.2 The collapse of the distinction

In ‘Truth’, Austin says that ‘many utterances which have been taken to be statements...are not in fact descriptive, nor susceptible of being true or false.’ He gives some examples, performatives being one. However, in the same paper, he states that it is common for statements to have a performatory aspect. He explains that ‘[I]t is common for quite ordinary statements to have a performatory ’aspect’: to say that you are a cuckold may be to insult you, but it is also and at the same time to make a statement which is true or false.’ The utterance ‘you are a cuckold’ is both: it is performative, to insult you, and it is descriptive, it is a statement, which is either true or false.

The difficulty is that this position seems inconsistent: on the one hand Austin seems to be denying performatives the capability to indicate truth or falsehood, but, on the other, he seems to grant them this ability. As a result, the fundamental distinction between performatives and constatives seems to be threatened, and Austin himself quickly realises this.

In particular, he recognises that for both kinds of utterances we often appraise the relation between the words and the world in the same way, using the same family of terms which belong to the dimension of ‘truth’/ ‘happiness’. Any utterance is appraised in relation to both the appropriate circumstances under which it is uttered, and the facts which the utterance somehow ‘corresponds’ to. Thus, constatives are assessed (being true or false) in relation to facts; however, the ‘happiness’ of performatives are assessed in relation to facts: we estimate rightly or wrongly; we find correctly or incorrectly; we argue soundly; we advise well; we judge fairly; we blame justifiably. In all these cases, our assessment relies on the facts: ‘the question always arises whether the praise, blame, or congratulation was merited or unmerited’. Equally, ‘such adverbs as ‘rightly’, ‘wrongly’, ‘correctly’, and ‘incorrectly’ are used with statements too.’ All this make us question the original distinction between two kinds of utterances, constatives which are merely true or false and correspond to facts, and performatives, which were thought not to be true or false in virtue of neither describing nor stating things, and therefore not corresponding to facts. As a result, Austin asks ‘Can we be sure that stating truly is a different class of assessment from arguing soundly, advising well, judging fairly, and blaming justifiably? Do these [performatives] not have something to do in complicated ways with facts?’ In assessing a performative to be happy or unhappy, using the adjectives above, ‘[F]acts come in as our knowledge or opinion about
In other words, the happiness or unhappiness of performatives, which originally were thought to be independent of correspondence to the facts, turns out to be related to facts, as are constatives.

A similar difficulty arises when we consider constatives, whose truth values were originally thought to be independent of the circumstances of uttering the words. Austin gives the following example. ‘Suppose that we confront ‘France is hexagonal’ with the facts, in this case, I suppose, with France, is it true or false? Well, if you like, up to a point; of course I can see what you mean by saying that it is true for certain intents and purposes.’ According to Austin, it is a ‘rough description’. But we can’t simply assess if it is true or false. Thus, as we saw earlier, ‘it is good enough for a general top-ranking general, but not for a geographer’. It is difficult to see how we can say it is true or false, without taking the circumstances of uttering it into account. Take another example: ‘Lord Reaglan won the battle of Alma’. This is good enough for a school book, but not for a historical research. More examples from ‘Truth’ include: ‘Belfast is north of London’, and ‘the galaxy is the shape of fried egg’ which we also discussed above. In all these cases, it seems that we can’t tell if the statement is true or false without taking into account the circumstances under which it was uttered.

The upshot of this is that the distinction between performatives and constatives collapses. The distinction was supposed to show us that we have, on one hand, utterances which are true or false, which correspond to the facts and are independent of the circumstances of utterance, and, on the other hand, utterances which are not true or false, and are assessed according to the circumstances under which they are uttered. The above examination shows us that both kinds of utterances are often related both to facts and to the circumstances under which they uttered, and that they are both assessed in similar ways. And the key reason for this, according to Austin, is his view of ‘truth’ as a dimension-word. The terms which we use in assessing the performatives overlap with the terms we use in assessing constatives: we use the same family of words to describe and assess both performatives and constatives. Austin concludes ‘[W]hen a constative is confronted with the facts, we in fact appraise it in ways involving the employment of a vast array of terms which overlap with those that we use in the appraisal of performatives.’

The constative/performative distinction is instable; it distinguishes between constative: saying something, which is either true or false, and performative: doing something, which is either happy or unhappy. Now we have seen that the true/false and happy/unhappy criterion is flawed.
Next, we will see how Austin introduces a new theory of speech act to replace the collapsed distinction, how it is related to truth, and how the saying/doing distinction is also flawed.

4 Austin’s theory of speech acts and the role of truth into it

The failure of the distinction between ‘performatives’ and ‘constatives’ prompted Austin to propose a new theory of speech acts. His key idea was that in analysing utterances we should distinguish between a locutionary act and an illocutionary act.

The locutionary act is ‘the utterance of certain noises, the utterance of certain words in a certain construction, and the utterance of them with a certain ’meaning’…’43 This contrasts with the illocutionary act. As Austin puts it: ‘To determine what illocutionary act is so performed we must determine in what way we are using the locution:

[asking or answering a question,
giving some information or an assurance or a warning,
announcing a verdict or an intention,
pronouncing a sentence,
making an appointment or an appeal or a criticism,
making an identification or giving a description,
and the numerous like.44

What we have here, then, is a new theory which distinguishes between two acts, and between two different levels of meaning in accordance with the two acts. Thus, for example, we distinguish between the meaning of the utterance: ‘Shoot her!’, and the force of that utterance, which depends on the circumstances but could consist in urging, or advising, or ordering me to shoot her.45 The words uttered in the locutionary act have a ‘meaning’, or a ‘locutionary meaning’, and in the ‘illocutionary act’ they have a ‘force’. Note that Austin gives a technical sense to both ‘meaning’ and ‘force’. ‘Admittedly we can use ’meaning’ also with reference to illocutionary force- ’He meant it as an order’, etc. But I want to distinguish force and meaning.’46

Before going back to Austin’s views on literal meaning, let me clarify one important point on the collapse of the constative/performative distinction: In secondary literature, the collapse of the distinction is seen, mainly, as a result of Austin’s finding that stating is an act, just like performatives: stating, in the new theory, is just one of the illocutionary
acts. For example, John Searle writes that ‘one of Austin’s most important discoveries, [is] the discovery that constatives are illocutionary acts as well as performatives, or, in short, the discovery that statements are speech acts.’ Austin is clear enough on this point: ‘Stating, describing, etc., are just two names among a very great many others for illocutionary acts.’

The constatives/performative distinction is thus flawed, because it is based on a distinction between saying something and doing something, and Austin finds that constatives and performatives are both ‘doing’ and ‘saying’. The new theory gives him a sophisticated tool to clarify how each utterances consists of different doings (acts), in different ways.

My discussion in 3.2 above on the collapse of the distinction does not exclude the importance of this ‘discovery’, but it focuses on a neglected aspect of Austin’s work: his claim that ‘true’ is a dimension-word.

Now we can go back to the Hansen-Crary debate, and clarify the role of ‘truth’ in the new theory of speech act. Crary denies that sentences have a fixed meaning; Hansen affirms that sentences have a fixed meaning, and that this fixed meaning suffices to determine whether, independent of occasions of use, the sentences can be used to say something true or false. My claim is that for Austin, contrary to Crary’s claim, there is an invariant meaning the sentences carries through different uses, this is the locutionary meaning. However, contrary to Hansen’s claim, this meaning is not to be identified with what is true or false: what is true or false is the speech act as a whole and not any one part of it.

Let us start with the first of these two points. If we examine Austin’s text we will see that he is happy to say that there is an invariant meaning the sentence carry in all its uses, and that this is the locutionary meaning, as Hansen argues.

Austin’s position runs something like this. On the one hand, as Crary argues, Austin stresses that the illocutionary and locutionary acts are inseparable as acts. Austin writes of the relation between them: ‘in general the locutionary act as much as the illocutionary is an abstraction only: every genuine speech act is both…’ Again, he says that ‘[T]o perform a locutionary act is in general, we may say, also and eo ipso to perform an illocutionary act.’ It seems that this is what impresses Crary and gives strength to her reading. However, on the other hand, it seems clear that Austin doesn’t preclude the idea that the same locutionary meaning may persist across different forces. I agree with Hansen that it doesn’t seem that there is textual evidence for taking Austin to reject such an idea.
I therefore agree with Hansen that for Austin it is possible to abstract from the whole speech act ‘meaning’ and ‘force’, and that it is possible to abstract the ‘literal meaning’ of sentences, the ‘locutionary meaning’ from different speech acts. For example, by uttering ‘Shoot her!’, we might have different forces, such as urging, or advising, or ordering me to shoot her, but we have one locutionary meaning ‘shoot here’. In that sense, there might be literal meaning which is invariant in different uses. However, Hansen conflates meaning and truth evaluability in making his claim that the literal meaning, once abstracted, is what is true or false. On this point, I think Austin is better interpreted as follows.

For Austin, it is not the locutionary meaning which is to be identified with what is true or false. Rather, it is the whole speech act. From my earlier explanation of truth as a dimension-word and the collapse of the constative-performative distinction, we can now explain what is true or false for Austin.

Austin writes, commenting on the relation between the old distinction and the new theory of speech acts:

With the constative utterance, we abstract from the illocutionary . . . aspects of the speech act, and we concentrate on the locutionary. . . With the performative utterance, we attend as much as possible to the illocutionary force of the utterance, and abstract from the dimension of correspondence with facts. . . Perhaps neither of these abstractions [constative as focusing on the locutionary, and performative as focusing on the illocutionary] is so very expedient: perhaps we have here not really two poles, but rather an historical development. Now in certain cases, perhaps with mathematical formulas in physics books as examples of constatives, or with the issuing of simple executive orders or the giving of simple names, say, as examples of performatives, we approximate in real life to finding such things. It was examples of this kind, like ’I apologize’, and ’The cat is on the mat’, said for no conceivable reason, extreme marginal cases, that gave rise to the idea of two distinct utterances.

Austin here makes explicit the instability of the distinction between constatives and performatives that I identified above. Whilst there are extreme cases where the distinction is clear, the vast majority of utterances fail to conform to such a division. It is on the basis of this realisation that Austin wants to introduce his new theory of speech acts, the collapse of
the old distinction having been driven, as we saw earlier, by the recognition that ‘true’ is a dimension-word, and that we use the same family of words to appraise both kinds of utterances. In particular, it seems that Austin recognises that at the heart of the collapse of the distinction is the realisation that we can’t separate the two categories by appealing to two distinct notions of appraisal: true/false and happy/unhappy.

Now the question that arises is what is true or false in the theory of speech acts?

I claim that the most plausible reading of Austin’s position is that being true or false is to be assessed in relation to a whole speech act, and not any part of it. This position is consistent with the moral of the collapse of the first distinction, where we had to take into account that the terms of assessments of ‘true’ and ‘false’ merged and overlapped with terms of assessments of ‘happy’ and ‘unhappy’. The lesson there was that both types of assessment generally depended on, and were determined by, both facts and circumstances of utterance. As Austin points out, it is only in extreme cases, such as propositions of logic and mathematics, and performatives such as ‘I apologize’, that we have constatives which are true or false regardless of the circumstances, and performatives that are happy or unhappy regardless of the facts. Being true or false in the general run of utterances, then, is to be understood being determined by both circumstances and facts. As we have seen, the different dimensions of being ‘true’ or ‘false’ are related to both of these features, which are present in the different terms we typically use to appraise utterances.

The problem with identifying the locutionary meaning as that which is true or false is that it treats it as a ‘proposition’ which is to be appraised as true or false regardless of the circumstances under which it is uttered. Whilst I take Austin to agree that the locutionary meaning might be shared by different speech acts and that it is something that we abstract from those different speech acts, I don’t take Austin to identify the locutionary meaning with what is true or false. If the locutionary meaning, which we abstract from different speech acts, were by itself and independent of being uttered under specific circumstances, since it is abstracted from the actual circumstances under which it is uttered – true or false, then being true or false would not be related to the circumstances under which the words are uttered. This is precisely the opposite of what I have tried to show for Austin: that the circumstances under which we utter the words are vital for applying the terms of the ‘truth’ family. Whilst, in extreme cases, the abstract component of the dimension word can be used on its own without reference to the circumstances
of use, in almost every normal case the abstract element is too weak to be used and, instead, other words in the same family are employed, words that better reflect the context.

It is therefore my position that it is the whole speech act which is to be judged ‘true’ or ‘false’, and not the locutionary meaning: being true or false is related to issuing a whole speech act in a specific situation.

One last clarification: We said that we distinguish between sentence-meaning and word-meaning. Austin’s views of sentence-meaning mirrors to a large degree his position on word-meaning. Whilst the two accounts are distinct, what they share is a caution concerning the attempt to isolate a common feature, either representing the essential factor in the specification of a concept or the essential component of sentence-meaning, independent of the influence of context. In both cases, Austin recognises the initial attraction of such an approach, and is happy to concede that there may be a common feature shared by the discussed concepts on occasion, and an invariant literal meaning for sentences. However, his key point is that we need to take into account, in analysing ‘meaning’, the different contexts and circumstances in which the word is used, or the sentence uttered. Nevertheless, the two accounts are different. As we said above, Austin studies three kinds of words which have multiple meanings, ‘trouser-words’, ‘adjuster-words’ and ‘dimension-words’. In addition, he concedes that some words have one and the same meaning in all their uses. There doesn’t seem to be parallel distinctions when it comes to sentence-meaning, and there is no reason to expect such a parallel. We don’t have space here to go into details on this. But we need to guard against an unnecessary generalization or confusion in the relation between sentence-meaning and word-meaning.

In this paper, I have argued that the discussion on word-meaning, in this case the word ‘true’, is crucial to understand Austin’s views of sentence-meaning. Austin’s claim that ‘true’ is a dimension-word serves him for, first, arguing for a collapse of the performative/constative distinction, and, then, for identifying what is ‘true’ or ‘false’ with the issuing of the whole speech act in a specific situation.

The conclusions of the debate on sentence-meaning might be summarised as follows: contrary to Crary’s claim, there is an invariant meaning, a literal meaning, that sentences carry through different uses, this is the locutionary meaning. This ‘literal meaning’ is context-independent. However, contrary to Hansen’s claim, this meaning is not to be identified with what is true or false: what is true or false is the speech act as a whole
and not any one part of it, or more accurately, the issuing of the speech act in a specific situation. Being ‘true’ or ‘false’ is context-dependent.

5 Conclusions

In summary, I have argued that there is a middle ground that should be taken between Crary’s and Hansen’s readings of Austin’s views on literal meaning. According to my analysis, his claim that ‘true’ is a dimension-word plays a crucial role in his theory of speech acts. Specifically, ‘true’, as a dimension-word, has one and the same semantic function in all its uses, but different specific meanings according to the context, a finding that causes Austin to abandon the performative/constative distinction.

It is only against this background that we come to appreciate the role of ‘truth’ in his theory of speech acts: for Austin, (at least some) non-ambiguous sentences have a fixed meaning, which they carry across all occasions of use. Such sentences meaning what they do, however, does not suffice to determine whether, independent of these occasions, the sentences can be used to say something true or false. What is true or false is the whole speech act, and not any part of it.

Notes

1 [8], 59-60.
2 [8], 60.
3 [8], 68.
4 [8], 69-70.
5 [12], 3.
6 [12], 6.
7 [3], 98.
8 See [12], 6-7. He is aware that the text is not conclusive, which is why he only thinks we might attribute to Austin the literal meaning view.
9 See [12], 6.
10 [19], 110.
11 [1], 110–11.
12 Travis mentions the idea that true is a dimension-word in [20]. However, even when he does mention it, it doesn’t play the role in his work I claim it plays in Austin’s work. Furthermore, if my claim is correct, there will be some considerable differences between Austin and Travis on truth. It is, however, beyond the scope of this paper to pursue these differences further. The same is applied to Francois Recanati’s work. Although my interpretation is compatible with his, he
doesn’t give the claim that ‘true’ is a dimension-word the crucial role I give. See [15] and [16]).

13 [1], 64.

14 ‘Yellow’ and ‘horse’ are Austin’s examples; ‘bank’ is mine.

15 [4], 74. Austin takes Plato to be committed to the search for one meaning for each word. Elsewhere, he also contrasts this with Aristotle.

16 The other types are trouser-words and adjuster-words.

17 [1], 71.

18 Austin gives other examples in his writings, such as ‘good’, and ‘freedom’.

19 [4], 250-251.

20 Austin discusses the different terms of the family we use on ordinary language in ‘Performative Utterance’, see [3], 142-147 and [4], 129-130 and 250.

21 [4], 130.

22 [1], 71.

23 [1], 71.

24 As we use them to assess utterances, only. Note that words have other uses. ‘True’, for example, is one word in the family of ‘real’. The relation between the different uses of words is a complicated issue, and Austin doesn’t say a lot about it.

25 Austin discusses this ‘descriptive fallacy’ in a number of different places: see [4], 97-103, 130-132 and 233-234; and [3], 1-4 and 100.

26 Austin in How to Do Things with Words was suspicious of the two terms: ‘descriptive fallacy’ and ‘statements’, ‘perhaps this is not a good name, as ‘descriptive’ itself is special. Not all true or false statements are descriptions, and for this reason I prefer to use the word ’Constative’.’ [3], 3.

27 [4], 131.

28 [4], 131.

29 In the three later works mentioned above, Austin examines the distinction before declaring that it is not working. Most of what follows depends on the characterization of the distinction as it appears in the major work, How to Do Things with Words.

30 [3], 6.

31 [4], 236.

32 [3], 133.

33 [4], 131.

34 [4], 133.

35 [4], 133.

36 [3], 141.

37 [3], 141.

38 [3], 142.

39 [3], 142.

40 [3], 143.
41 [3], 143.
42 [3], 143-144.
43 [3], 94.
44 [3], 98-99.
45 The example is from Austin, *How to Do Things with Words*, 101-102.
46 [3], 100. The distinction between the two acts, and between meaning and force, has been interpreted in different ways. For example, John Searle ([17]) and ([18]) and Jennifer Hornsby ([13]) and ([14]) found it unclear in some points, and suggested different ways to improve it. I don’t claim that the distinction is clear cut, but I think that we don’t need, for the purposes of this paper, to go into complications regarding how to demarcate the illocution exactly.
47 [18], 157.
48 [3], 148-149.
49 We don’t need here to go into details of how to distinguish exactly between the different acts.
50 Similar readings are offered by other interpreters, Max Black ([6]) and Jennifer Hornsby ([13] and [14]), for example. All these readings, correctly, point out that the saying/doing criterion is flawed, and they point out that the true/false and happy/unhappy criterion is problematic, in different ways. However, none of them brings out the importance of ‘true’ as a dimension-word.
51 [3], 147. He adds, ‘But, of course, typically we distinguish different abstracted ‘acts’…’
52 [3], 98.
53 [3], 101-102.
54 [3], 146.
References


New Experiments On Psychological Egoism

Nenad Natan Popovic

Abstract

Psychological egoism is the doctrine that all of our actions are ultimately motivated by self-interest, even including seemingly altruistic actions. Although this view is commonly referred to as “vacuous” or “trivially true” because of its immunity to counterexamples, I argue that there are possible observations that can refute this thesis. I describe the experiments, identify the results that could in principle falsify the strongest possible version of psychological egoism and formulate a probabilistic argument against this view. Ultimately, I conclude that a “trivially true” formulation of psychological egoism is an empirical theory that is in all likelihood false.

Keywords: psychological egoism, altruism, self-interest, unselfish behavior

1 Introduction

Anyone who has ever taught a philosophy course that includes the doctrine of psychological egoism can easily imagine the following dialogue taking place in the classroom. In fact, the excerpt below is a brief description of an exchange of arguments that many of us have witnessed numerous times. Let (DEF) stand for a defender of psychological egoism, and let (OPP) stand for an opponent of this view:

(OPP): Psychological egoism must be false. We don’t always act out of self-interest. There are many people who donate their time, effort and money to helping others.

(DEF): No, that does not show that they are not acting out of their self-interest. If you volunteer or give money to charity, perhaps
you are doing it for tax purposes or to have something to put on your resume; perhaps you just want to feel good about yourself, or perhaps you know that you would feel bad if you didn’t help; maybe you want others to think highly of you because you are helping people in need; or maybe you just can’t stand to see others suffer, so by helping them you are helping yourself. I can’t tell which explanation is true, but it must be one of these. So, people who help or care about others are still ultimately motivated by their self-interest.

(OPP): What about people who risk their own lives or their own well-being to save others? What about a parent who jumps in front of a car to save their child? A soldier who throws himself on a grenade to save others?

(DEF): Perhaps they know that they couldn’t live with the fact that they didn’t try to do something, or maybe they identify their self-interest with the well-being of others. Either way, it still does not mean that they are not motivated by self-interest. In those circumstances people simply choose what they think is the lesser of the two evils at the time when they are acting.

(OPP): What if I just give you my wallet and tell you to keep the money. Believe me, I don’t want to do that at all. But if I did it just to prove a point, would that count as behavior contrary to psychological egoism?

(DEF): No, it wouldn’t. If you did it because you wanted to show that psychological egoism is false, then that would be your self-interest and you would still be acting in accordance with this view.

(OPP): So there couldn’t possibly be a case in which a person doesn’t act out of their self-interest, then. If a person did help others and thereby promoted his own welfare, then he is self-interested, while if in the same circumstances he helped others but sacrificed his welfare to do so, then he is equally self-interested. If that is what psychological egoism is stating, then the theory is vacuous. You are just defining actions in terms of self-interest, and self-interest itself is defined so broadly that absolutely anything can be categorized under it. Of course there can’t be a counterexample to this theory.
While it is possible to distinguish between several different formulations of psychological egoism,\(^1\) my focus will be on a version of this view that has been implicitly referred to in the previous dialogue. Thus, by psychological egoism I shall understand the view that all of our actions are ultimately motivated by self-interest, and that even seemingly altruistic motives are in fact just instances of motives intended to further our own self-interest. More specifically, my target will be the following thesis:

\((PE)\): For every agent S and for every action A, S will willingly perform A at t0 only if S believes at t0 that performing A will promote S’s self-interest.

Psychological egoism is notoriously difficult to disprove if the notion of “self-interest” is understood in the way in which it was used by (DEF) in the previous dialogue.\(^2\) As it has been illustrated, no matter what kind of counterexample we try to offer, it seems that there is always an explanation (more accurately, a disjunction of possible explanations) available to the supporter of (PE). It looks as if this theory is therefore immune to any scenario and that it cannot be refuted by means of any conceivable counterexamples. Perhaps this point is best summarized by Feinberg:

‘Until we know what they [psychological egoists] would count as unselfish behavior, we can’t very well know what they mean when they say that all voluntary behavior is selfish. And at this point we may suspect that they are holding their theory in a “privileged position”—that of immunity to evidence, that they would allow no conceivable behavior to count as evidence against it. What they say then, if true, must be true in virtue of the way they define—or re-define—the word “selfish.” And in that case, it cannot be an empirical hypothesis.’\(^3\)

I take it that this is the primary reason why this version of psychological egoism is commonly referred to as a “trivially true,” “vacuous,” “uninteresting”, “unfalsifiable in principle” or just “trivial.” This is also the primary reason why most of the literature on this subject instead focuses on other, “more interesting” formulations of psychological egoism.\(^4\)
I will argue that, contrary to this common misconception, there are possible observations that can refute (PE). I will specify some experiments that can be performed in order to test this theory empirically and identify the outcomes that would show that the thesis under discussion is false. Once this result is established, it will become apparent that psychological egoism cannot be a trivial thesis, even when it is defined in a way which includes seemingly altruistic acts. Rather, my ultimate conclusion will be that (PE) is false in all likelihood, although we may have to wait for the experiments to be carried out in order to get final evidence.

2 The (PE) Experiments

Suppose that a group of subjects is invited to participate in a study involving the use of polygraphs. Each participant will receive an honorarium for answering a number of questions administered by a polygraph specialist. Suppose also that the subjects who have been chosen are not familiar with the techniques used in this process.

For any subject S, a set of irrelevant questions and control questions is administered first. After this phase has been completed, the subjects are then asked if they would like to donate a small amount of their honorarium to some charitable organization. The subjects are told that they are under no obligation of any kind to make a donation, nor that it is expected of them to do so in any sense. If S refuses to make a donation, the test is over and no further questions are asked. If, on the other hand, S decides to make a donation, the test continues. S is then told to confirm the donation by touching the screen of a monitor several times, until the final confirmation window appears informing S that his donation will be processed as soon as he presses a green “YES” button on the screen. However, this button is not available and can only be enabled by the specialist administering the test. At this time, S is presented with a set of relevant questions that pertain directly to S’s possible motives for making a donation. This set of questions consists of claims that would usually be used by a supporter of psychological egoism, and S is supposed to provide an answer to each of them. Suppose that the set of relevant questions contains the following:

Q1: Do you want to make this donation because you are expecting some benefit from it?
Q2: Do you want to make this donation because you want to feel good about yourself?

Q3: Do you want to make this donation because you believe you will get any pleasure out of making this donation?

Q4: Do you want to make this donation because you would feel bad if you did not do it?

Q5: Do you want to make this donation because you believe that you would regret it later if you did not do it?

Q6: Do want to make this donation because you want to impress the test administrator in some sense?

Q7: Do you want to make this donation because you want to impress someone else later on when you tell this story?

Q8: Do you believe that something negative will happen to you if you do not make a donation?

Q9: Do you believe that something positive will happen to you if you make this donation?

Q10: Is it in your self-interest to any degree to make this donation?

Suppose that the test administrator is instructed to enable the button that allows S to make a donation if and only if S gives a negative answer to all control questions Q1 – Q10 and passes the polygraph test for each of those questions. If both of these conditions are satisfied, the administrator will immediately enable the confirm button and the donation will be processed as soon as S touches the screen. If S gives an affirmative answer to any of the control questions Q1 – Q10, the administrator will not enable the confirm button but will inform S that there was a problem with the software and that S’s donation cannot be processed at this time. Similarly, S will not be in a position to make a donation if he fails the polygraph test for any of the control questions that he provided a negative answer to.

If this experiment is performed and S ends up making a donation, we will be provided with evidence that (PE) is false since S has probably acted in a purely altruistic manner. If several subjects end up making a donation the evidence will get proportionally stronger, and if a significant portion of all tested subjects passes the test (PE) will be outright refuted.6
We have thus identified outcomes that would show that (PE) is false, which means that psychological egoism is actually an empirical thesis that can be tested in principle by means of experimentation. While we may have to wait for the results of those experiments to come in before we can proclaim our final verdict, there are at least some indicators that we should expect (PE) to be refuted by those results. If (PE) were true, then it is highly improbable that anyone in the world could ever truthfully give a negative answer to all of the questions (Q1) – (Q10). Yet, there is something very peculiar about this result since we can undoubtedly envision that occurring, which already furnishes us with a strong reason to suspect that (PE) is in fact false. Also, there clearly are numerous individuals who believe that genuinely altruistic acts are possible, and it seems reasonable to expect that they would truthfully deny any questions suggesting the contrary. This also indicates that a significant portion of the subjects should be expected to pass the test and thus provide us with relatively strong evidence against (PE).

3 The Ultimate Objection

The experiments that have been described provide opponents of (PE) with a threatening argument against this doctrine. We can thus easily imagine the initial dialogue between (OPP) and (DEF) extend in this direction:

(OPP): What about people who genuinely and sincerely believe that their actions are not motivated by self-interest? What if we conduct experiments using polygraph and it turns out that a significant number of people can pass this test? Would that not constitute ultimate evidence against psychological egoism?

At this point, a supporter of (PE) seems to be left with only two options. He can either concede that the aforementioned experiments are relevant and thus expose the doctrine he is defending to refutation, or he can resort to the following argument:

(DEF): Our motives for acting are not infallibly known. Just because people believe that they are not acting selfishly, it does not follow that they are not acting selfishly. Perhaps they are motivated by subconscious mechanisms and are not aware of their deep egoistic motives; perhaps they are self-deceived or simply mistaken about their real motives. Either way, the experiments
can only show that some people may believe that they aren’t acting out of self-interest, but they can’t demonstrate that people aren’t acting out of self interest. Therefore, the experiments described are irrelevant. They can’t possibly show that psychological egoism is false.

Is the reply given by (DEF) satisfactory and can we simply dismiss the previous argument against (PE) on the grounds that people can be unaware or mistaken about their own reasons for acting? Should we conclude that the experiments are altogether irrelevant to the status of (PE) since we may not infallibly identify the factors that motivate our actions? I will argue that the argument given by (DEF) is fundamentally flawed and that, although the main premise is true and our motives for acting are not infallibly known, the conclusion does not follow.

4 The Ultimate Reply and the Intuitive Appeal of the Probabilistic Argument

As mentioned previously, it would be outright unreasonable to deny the first premise of the argument. Of course it is possible that people can be mistaken about their own motives. However, it is prima facie implausible to claim that all people who ever pass the lie detector test must hold false beliefs about their own motives. Yet, this is exactly what (DEF) has to maintain in order to dismiss the experiments.

So, while this experiment does not rule out the possibility of subjects being mistaken about their own motives and while it is a far cry from concluding that all subjects who pass this test are de facto motivated by unselfish reasons, we must be reminded that it is the supporters of (PE) that are arguing for a universal thesis. To falsify (PE) we only need one instance of behavior that is not motivated by self-interest, and the argument simply states it is improbable that all of those who pass the lie detector in the circumstances described are mistaken about their own motives.

Additionally, it seems intuitively clear that the results of the experiments cannot be irrelevant. To illustrate this, suppose that out of a group of a thousand subjects only one ends up making a donation. In this case it would not seem as unusual to argue that a small percentage of individuals can be mistaken about what really motivates their actions. However, suppose that the experiments show that more than 50% of people who are tested in fact end up in that category. At this
point, making the same claim as before would already start to sound farfetched, as we would have to maintain that most people are either self-deceived or mistaken about their own motives in order to preserve (PE).

Even worse for the supporter of (PE), this argument could be pushed further. Suppose that altruistic ideas become culturally dominant at some time in the future, to the point that nearly 100% of the population believes that we should help others even if it does not advance our own self-interest, and that nearly everyone actually acts on that principle most of the time. In that case it would not be hard to imagine that nearly 100% of the tested subjects would pass the polygraph test. Yet, supporters of (PE) would still have to attribute false beliefs to almost everyone and argue that virtually all humans are systematically mistaken about their motives with respect to most of their actions. At least, this should be considered a highly undesirable consequence of (PE).

In conclusion, it seems clear that our intuitions may at least somewhat fluctuate depending on the number of people who end up passing the polygraph test. There is an intuitive difference between no one ending up in this category, versus someone or nearly everyone honestly believing that their actions were not motivated by self-interest. It seems clear that the actual results of the experiments bear at least some relevance, and therefore the experiments and the probabilistic argument against (PE) cannot be entirely dismissed based on the claim that our reasons for acting are not infallibly known. So, although the main premise of the argument made by (DEF) is true, its conclusion does not follow.

Additionally, the case against the argument made by (DEF) can be further strengthened by focusing on the circumstances of the subject. According to the objection, subjects who pass the polygraph test must hold false beliefs about their own motives, presumably because they are mistaken, self-deceived or unaware of their deep egoistic motives.

However, we must be reminded that the setup of the experiments includes questions that have been specifically designed to address these concerns. According to the setup, subject S has been explicitly asked about possible egoistic motives. Subject S has thus been placed in the position to be reflective about his true reasons for acting as he considers those questions and how they relate to his actions. If S was mistaken or failed to even consider any alleged egoistic motive that he might have while he was performing the action, he surely became aware of any such motive during this part of the experiment and reflecting on the questions
will increase the likelihood of identifying its potential relevance.

Perhaps this may not be enough for the subject to uncover just about any deep egoistic motive that he might have, but it certainly minimizes the likelihood that all subjects who pass the test hold false beliefs about their motives. Besides, the setup of the experiment can easily be further amended to neutralize the force of even this objection. For example, we can stipulate that the subjects must undergo a session with an expert who specializes in helping the subjects uncover their deep or hidden motives before answering the questions. By virtue of this setup that can be even further amended if needed, it is prima facie unlikely that all subjects who end up giving negative answers are thus motivated by hidden reasons that are inaccessible to them.

Furthermore, the polygraph test itself is designed to detect any unsettling pattern that the subject might exhibit. As such, it should be particularly troubling for people who are sensitive or worried about a certain subject or self-deceived about their real motives. Consider someone who would find it painful to know that what he sees as purely altruistic motives are in fact just instances of egoistic motives, and yet who is by the nature of the experiment suddenly placed in the position to directly process this possibility. It seems fairly likely that it would be detected during the test. As we have seen previously, rather than being mistakenly classified as false negatives, subjects who fall in any such category are likely to produce a disproportional number of false positives.

In summary, if subject S is wrong about his own motives, it is either because S is unaware, mistaken or self-deceived about his real motives. If S is unaware or mistaken about his true motives, the questions will likely make him aware of it; if S is self-deceived or refuses to acknowledge his true motives, the polygraph test will likely detect it. Therefore, it is highly improbable that all subjects who pass the test are wrong about their own motives. Therefore, it is highly likely that (PE) is false.

The previous argument tried to establish that the individuals who pass the test are unlikely to be mistaken about their motives, but it is important to note that the argument is fatal to (PE) practically as long as there is any non-zero probability that any individual subject who passes the test is speaking truly. For example, suppose that each individual subject who passes the test is 90% likely to be mistaken. By simple probabilistic calculations, if 7 subjects pass the test on this assumption it is already more than 50% probable that at least one of them is actually speaking truly; if 22 subjects pass the test the probability of (PE) being false is already above 90%. So, even if we disregard the
previous points and allow for the sake of the argument that subjects are likely to be self-deceived or otherwise mistaken about their own motives, even a relatively small number of confirmations will still allow us to conclude that it is highly improbable that (PE) is true. The only way for a supporter of (PE) to escape this result is to maintain that it is impossible for the subjects to speak truly in those circumstances, or if the experimental results actually fail to produce sufficient evidence.

In conclusion, it cannot be denied that our reasons for acting are not infallibly known. This fact, however, does not constitute sufficient ground for dismissing the probabilistic argument against (PE) and the experiments intended to provide the evidence. There is a clear intuitive appeal that the results of the experiments cannot be completely irrelevant, and it is highly improbable that all subjects who pass the polygraph test must hold false beliefs. The premise of the argument provided by (DEF) is true, but the conclusion simply does not follow. Consequently, (PE) turns out to be an empirical claim that can be tested by means of observation-based experiments rather than being a “vacuous” thesis. Far away from being “trivially true,” psychological egoism in the end is just a “sitting duck” waiting to be demonstrated false once and for all.

Notes

1 For a summary of possible interpretations of psychological egoism see [8, pp. 501-503].

2 [1] argues that the experiments done in the field of social psychology provide empirical evidence against psychological egoism and for the empathy-altruism hypothesis. See [11], [12], and [5] for criticisms of Batson’s conclusions; ([1]) also see [2] for replies. [11] argue against psychological egoism based on evolutionary theory. For criticism of Sober and Wilson ([11]) see [12]. For traditional philosophical arguments see [4], SermonXI and [7, p. 166]. While these arguments have merits that have widely been recognized and while they are effective against many formulations of psychological egoism, a defender of (PE) can still insist on taking the same route as described in the initial dialogue and argue that no genuine counterexample to the theory has been presented.

3 [6, p. 503]

4 For the most part, this seems to apply to both critics and supporters of the view. For example, [9] makes a distinction between weak and strong psychological egoism and argues that the former is true while the latter is false, although neither formulation is identical to (PE). There are exceptions, though: [3] defends a version of psychological egoism that is very similar to (PE) on the grounds that it is the best explanation for human actions.

5 A defender of (PE) will surely argue that we can be mistaken about our own motivation and that our reasons for acting are not infallibly known. Although
it should be perfectly clear that I am offering a probabilistic argument against (PE), I will specifically address this objection in the next section.

According to a study conducted by the National Research Council ([10, p. 214]) for subjects untrained in countermeasures polygraph tests can discriminate lying from truth-telling at rates well above chance, though well below perfection. Also, while the National Research Council study ([10, Appendix I]) questions reliability of polygraph testing, notice that the main problem lies with producing a large number of false positives. So, far more truth-telling people test as deceptive and fail the test than the other way around. While this can be seen as a flaw of the method, it will actually significantly increase the overall value of evidence against (PE) obtained by means of our experiments. Polygraph testing may not be useful to identify a spy out of a group of thousand people since hundreds of innocents will also fail the test, but it will be very useful for our experiments since the test is likely to mistakenly classify a genuinely altruistic motive as egoistic. For example, if each one of a hundred subjects who are tested acts altruistically and if the assumed accuracy rate of the method is 80% we should only expect to get 80 confirmations of altruistic subjects or at best a few more. If fifty out of a hundred subjects act altruistically, we will likely get only 40 confirmations. If there are ten subjects who acted altruistically, we should expect to see 8 confirmations and so on. Given the two results established by the National Research Council, even a single case of altruistic behavior recorded by means of this experiment should count as very strong evidence against (PE).

Of course, there might be some overlooked selfish benefit not included in Q1 – Q10. If any such motive is identified, the questions should be adjusted accordingly.

References


Present Rights for Future Generations

Charlotte Unruh

Abstract

In this paper, I defend the view that within a rights-based ethical framework, the moral status of future generations is best understood as that of present rightsholders. I argue that in this way it can be justified that we have obligations towards future generations. This justification in turn is of great relevance for many issues in moral theory and applied ethics.

In the first part of the paper, I argue that the fact that future persons will have rights in the future cannot fully account for present obligations. The missing link in this argument cannot be provided by approaches that infer those obligations mediately.

In the second part of the paper, I argue that existing is not a necessary condition for being a rightsholder. First, our own future selves should be said to have rights even though they do not exist at present. Second, even at present, uncertainty challenges the relationship between rightsholders and obligation bearers: often enough, obligations depend on presuppositions or suspicions about other persons’ existence. In light of these cases, we should conceive of rightsholders as place holders, that is, sets of (actual) individuals whose existence or identity can be unknown or indeterminate, with specific properties. Therefore, future generations can coherently be said to have rights now that correspond to our present obligations towards them.

Keywords: future generations, rights, sustainability, nonexistence
1 Introduction

It is estimated that in 2050, global warming will have reached the two degree mark that is considered as critical for the conservation of current living conditions [12, p. 9]. Our carbon emissions, therefore, are contributing to potentially disastrous effects for future generations [11]. But do we thereby violate their rights?

Numerous challenges arise when we consider obligations towards future generations in a rights framework. Future generations do not exist yet. So how can we attribute rights to nonexistent beings? As a further complication, some of our actions determine who will come into existence in the future. This fact generates the so-called non-identity problem, since in those cases nobody will be made worse off by our actions, which contradicts the intuition that some such acts are wrong nonetheless [15].

In this paper, I concern myself with the first challenge. This is the problem of offering a consistent account of whether and in what sense rights can be ascribed to future people. Only on the basis of such an account we can tackle issues related to the non-identity problem and infer concrete implications for today’s policies and individual action.

The issue of sustainability and future generations is widely discussed in moral philosophy today. Several philosophers have proposed arguments, within a rights framework, of accounting for obligations towards future generations. These include what I call the direct approach (the argument that we can directly violate future rights of future people), the argument from commitment (which says that obligations towards future people are implied by our commitment to human rights or concepts such as dignity), and chain arguments (which can be understood as implying an intergenerational contract). All of these accounts infer present obligations from future generations’ future rights.

In the following, I will argue that this future-rights-for-future-persons-view faces a difficulty, namely that there is a missing link in the argument. I will then argue that future generations can, alternatively, be conceived of as present rightsholders. The present-rights-for-future-persons-view also captures our intuitions concerning future selves and cases that involve uncertainty.
2 A difficulty with inferring obligations towards future persons from future people’s future rights

The approaches that I will consider in the following all have in common that they do not accept the view that future generations have present rights. But they argue that we have present obligations towards them as future rightsholders. The first, direct approach consists in the argument that we can today violate future interests of future people, and thereby their rights. We therefore have an obligation to omit these acts. For example, Feinberg argues that “[t]he identity of [future generations] is now necessarily obscure, but the fact of their interest-ownership is crystal clear, and that is all that is necessary to certify the coherence of present talk about their rights” [8, p. 65]. A difficulty with this approach is that it is unclear how not yet existing rights can be binding for us right now. If we have obligations towards present people because of their present interests, it does not automatically follow that we also have obligations towards persons who will have interests or rights in the future. Those future rights cannot be violated today, since they do not exist yet. There seems to be a missing link in the argumentation.

The missing link in the argument might, one could argue, be provided by approaches that do not directly infer obligations from future persons’ rights. For example, Beyleveld, Düwell and Spahn argue that “[t]o deny that future humans have the same human rights as we do is to deny that there are human rights at all” [2, p. 550]. Committing ourselves to human rights implies caring for future generations, since we are in a position to violate their (future) rights [5]. Even though this argument can solve some of the problems of the direct approach because of its impartiality and independence from individuals, it does not provide the missing link from future persons’ rights to present duties. It is true that future humans will share important characteristics with us, and it is true that they will have rights, and it is also true that we can today act contrary to their interests and rights; but it is not clear why this should lead us today to act in certain ways. Since future rights do not exist at present, they cannot be violated at present. It is hardly reasonable to deny the existence of rights, but at the same time allow for their violation, that is, the existence of violated rights. But if we cannot violate future generations’ rights now, then an appeal to their rights cannot provide a reason for us to act in their favour.

Maybe such an account could be provided by a chain argument such as Howarth’s so-called chain of obligations. According to him, this chain is “mediated by the requirements of justice between contemporaries that
mandates our concern for future welfare. We owe it to our children, who will owe it to their children, who will owe it to their children, and so on as far as the mind can see. Thus our responsibility for the distant future follows directly from our obligation to our existing children” [13, p. 138]. It seems that this theory might be able to close the argumentative gap, since it relies partly on present rights towards present generations. But one could object that the problem is simply deferred. The idea is that one of the premises of Howarth’s argument, “[i]t is wrong to place others in a position where they are unable to fulfill their moral obligations to third parties” [13, p. 135], is not adequately justified in cases in which those third parties do not exist (yet). After all, on which basis should we accept that we have moral obligations to nonexistent third parties?

It seems reasonable to attribute future generations strong rights claims, especially in cases of actions that impose great burdens on future generations such as climate change. As Partridge notes, unlike mere beneficence, rights come with a stringency, urgency and a demand to be treated with respect [16, p. 43]. There are good reasons to think that we might need such a strong account in order to justify and enforce sustainable and future-oriented ways of acting. Even if we deny that the missing link is problematic, proponents of the account will probably have a hard time to account for a strong notion of future person’s rights. They cannot hold that those acts constitute a rights violation at the time they are taken (they will do so only in the future, since the rights that are violated do not exist at present). In contrast, rights claims of present persons can be violated at present. This qualitative difference seems to demand a substantially different consideration in moral theory.

I do not wish to claim that the missing link argument affects all theories that adhere to the future-rights-for-future-persons-view. There are many other theories that justify obligations towards future persons in a rights framework and that would need to be examined in this respect. Most recently, Bos’ chain of status connects the argument from commitment and chain arguments [3]. On top of that, there are theories that allow for the possibility that our obligation towards future generations does not correspond to a (future) right on their side. Such an account could, for example, be provided by reference to a Kantian duty of beneficence (although it is unclear whether these duties can be perfect) or so-called non-correlational obligations (which I believe would be quite problematic for other reasons as well2). Maybe those or other rights-based theories imply strong enough obligations towards future generations, without having to assume their present rights. As it has
hopefully become clear, though, the derivation of these duties can not be as straightforward as it might seem. Given that the future-rights-for-future-persons-view does have some difficulties, it might be worthwhile to take a look at alternative views as well.

Although I do not want to press the point here, another argument against the future-rights-for-future-persons-view is that it cannot account for the notion of past generations’ rights. For example, we could think of obligations to preserve the remembrance of the Holocaust as implied by corresponding rights of the victims of this horrible crime (though this might be not the only reason that implies this obligation). It therefore seems that past generations can also be said to have rights. Why should rights that have existed long time ago and will never exist again ground any moral rights claims to my present behaviour? It seems that one way to avoid this difficulty is to argue that both past and future generations have present rights.

3 Future generations as present rightsholders

The idea that future generations have present rights might not only sound implausible at first, it is also controversial whether it is actually possible. Among others, Beckerman and Pasek state that existing is a necessary condition for having rights [1]. This argument can be understood in two ways. First, it could be argued that a nonexistent person cannot be said to have anything at all. After all, it does not make sense to say that a nonexistent person has brown eyes or a blue backpack. This argument can be countered by pointing out that one does not have rights in the sense that one has those attributes. Rather, there are attributes that come with the moral status of being a rightsholder. Depending on one’s moral theory, this property could for example be rationality, autonomy, sentience, or a mixture of those. Based on this concept of rights, the argument might be explicated in a second way:

(1) If and only if some being $X$ has the property $A$, it is a rightsholder.

(2) To have properties (including property $A$), this being must exist.

(3) Future persons do not exist (yet).

\[ \therefore \] Therefore, future persons cannot have rights (yet).

In the following, I will refer to this argument as the nonexistence argument. Its conclusion contradicts the claim that future people can have
rights now. It therefore poses a problem for the view that future generations have present rights.

It might also be worth noting that the nonexistence argument can be employed contrary to its inventors’ intention. After all, if the nonexistence argument is accepted and if we add the rather uncontroversial premise

\[(4^*) \text{ Iff some being } X \text{ has the property } A, \text{ we can have obligations towards } X.\]

then it follows from \((2)-(4^*)\) that

\[\therefore \text{ Therefore, we cannot have obligations towards future persons (yet).}\]

It can be argued, therefore, that the views that infer present obligations from future rights can allow the nonexistence problem to sneak in through the backdoor. This is the case whenever a notion of duties or obligations is used that is justified by reference to the existence of other persons. This usually is the case for correlative rights and duties. If obligations can be characterized as other people’s rights claims, then the existence of the obligation, it seems, would depend on the existence of the rights claim. But also, if obligations are grounded in a notion of respect or fairness towards (autonomous, rational, sentient) beings, then it seems that these obligations do not exist in the case of future persons.

But it might be replied that even if this point is granted, it does not diminish the counterintuitiveness of a notion of rights that exist without a rightsholder. In the following, therefore, I argue that existing is not a necessary condition for being a rightsholder. I will present two arguments for that conclusion. The first argument is about our own future selves. According to the nonexistence argument, they could not be granted rights, which is implausible. The second argument is about epistemic uncertainty. I argue that if we grant rights in certain cases where future generations are not concerned, we commit ourselves to do so when they are.

### 3.1 Future Selves

Let me add two new premises to the nonexistence argument:

(4) Our future selves are beings that do not exist yet.

(5) Therefore, they cannot have properties (including property \(A\)).
It seems to follow that our future selves cannot be ascribed any rights (yet). But this seems to be an unacceptable conclusion. For example, in

_The bomb case:_ Tim hides a bomb under Tom’s desk just before his sabbatical, asserting that it will not explode before Tom will be back.\(^6\)

As Tom’s one year older self does not exist yet, it cannot have a right not to be harmed. But surely, we would insist that hiding the bomb constitutes a rights violation towards Tom. However, it could be objected that hiding the bomb violates his present rights not to be hurt in the future and therefore no future right needs to be assumed to condemn Tim’s action. This objection can be countered by pointing out that “present rights not to be hurt in the future” presuppose a morally relevant identity relation between present Tom and future Tom. While it is obvious that there is such a relation, it is worth putting in question how and under which conditions this relation is morally relevant. (I do not need to uphold the stronger claim that there is no way in which future people exist now.) This is especially so since one can ask, second, what morally relevant difference in status there is and should be between our future selves and persons not yet born. For example, consider

_The nursing home:_ Caroline, in the present moment (in 2016), is 10 years old. In 2096, she will, as we assume, become a nursing case. The same is true for Bert, who will be born in 2018.\(^7\)

Let us assume, as seems reasonable from a rights-based perspective, that there is a general right for nursing if it is needed. In 2096, undoubtedly both 90-year old Caroline and 78-year old Bert will have the right to be nursed. This is so although neither 90-year old Caroline nor 78-year old Bert exist yet. It seems absurd to argue that 90-year old Caroline, at the present moment, has a right to be nursed, but 78-year old Bert doesn’t. So either we concede that both of them have a right to be nursed (which could yield obligations for us now, for example, to build nursing homes), or we deny present rights to both future Caroline and Bert. Each way, there is no reason why we should consider future Caroline’s rights to be different from future Bert’s rights, nor does this seem intuitively plausible. The upshot is that future persons – at least those of whom we know that they will exist – have the same rights status as our own future selves.\(^8\)

However, conceding that neither our future selves nor future people have rights would lead to absurd conclusions. After all, this would mean
that rights only exist for split seconds, pertaining only to those persons who exist at that point in time. We could then not justify any moral decisions with reference to present rights, as the rights violated would always belong to the future beings affected by the decision. But this would mean that many of the ways in which we speak of rights are actually quite mistaken (e.g., “I have the right now that you do not hide this bomb”). If we want to avoid those difficulties, we have to reject the nonexistence argument.

3.2 The argument from uncertainty

How do we know that other people exist? While this is considered to be a somewhat tedious question in moral philosophy, it is also a very decisive one. If other people do not exist, according to the nonexistence argument, they cannot be ascribed any rights, and epistemic uncertainties therefore can affect our moral judgements. In the same way, they are influenced by ontological uncertainties about whether people do or will exist. Some rights are dependent on contingent properties. For example, you have a right to protect your property only with regard to things that you actually own. In contrast, basic human rights should be attributed to any human being. It has often been pointed out that the individual identity of rightsholders does not need to be known to ground obligations towards them – for example, the police officer has the duty to protect every citizen [1, p. 23]. In other words, epistemic uncertainty in those cases does not influence rights ascription. But what about ontological uncertainty? Consider the following scenarios:

*The river:* Imagine Bob is going for a walk at night along the riverside. Suddenly, he sees something moving on the water. It is not possible to go any closer, and he is not sure whether what he sees is really only the branch of a tree or a human being.

Does Bob have an obligation to call the water rescue? As long as there is a possibility that the thing in the water is really a human being, most of us would agree that he has this obligation – even if it turns out that Bob has only seen a branch. But, of course, there is a difference between cases in which a present rightsholder might exist and cases in which a future rightsholder might come into existence. Let us consider

*The playground:* Again, Bob is going for a walk at night along the riverside. He discovers that the fence of a nearby
playground is rusty. While it is still stable at the present moment, in a few years’ time, it might not be stable enough to prevent children from breaking through the fence and falling into the river.

The fence poses a danger to a child that might or might not exist today. If Bob has an obligation to repair the fence, he has that obligation with regard to a place holder that is indeterminate both in terms of existence and exact identity. I use the term “place holder” to define a set of actual individuals whose existence and / or identity is unknown or indetermined. Let us consider another case that involves ontological uncertainty:

The drug: A doctor has to decide whether to prescribe drug A or drug B to a severely ill patient. Both drugs will cure her disease. Drug A will cure it at once, while drug B requires a potentially stressful long-term treatment. However, drug A will have harmful effects to a fetus that is already existing or conceived within two weeks of taking drug A. Supposedly, it is not possible to rule out the possibility that the woman is already pregnant or going to conceive a child within two weeks.

The decisive question here is: Does it make a difference whether the woman is already pregnant or not when we decide whether to choose A or B? It is far more likely that our decision will depend on the probability that the woman will be pregnant in two weeks’ time, that is, whether a human being will be harmed by our decision or not.

What if we alter the drug case such that the doctor knows that the woman is not pregnant, but he knows that she will conceive a child in the next few days? Probably most of us would assume that he has a duty to take this fact into account, even though he knows for sure that no rightsholder exists at present. The point of the above examples is that ontological uncertainty, in some cases at least, does not affect our moral intuitions.

The distinction between uncertainty about present and future existence can only be gradual: It does not seem reasonable to accept the rights of a place holder that has only a 1% probability of present existence, and to categorically deny the rights of a place holder that has a 99% probability of existence in the near future.

If we accept these intuitions, we also have to accept that they blur the distinction between existent and non-existent beings with respect to
their status as a rightsholder, and that this difference might be better understood as gradual instead of dichotomous.

If I am right with the above argument, then it would be more adequate to define rightsholders in terms of place holders rather than “human beings” in premise (1) of the nonexistence argument, which would have to change to: If and only if some place holder X has the property A, it is a rightsholder. Since a place holder can have properties without existing, (2) would not apply any longer, and the conclusion would not follow.

3.3 A positive characterization

Ernest Partridge argues that future persons can have passive rights now [16, p. 60]; [7]. This sounds reasonable: future persons cannot act, but they can be treated in certain ways. Partridge illustrates his argument with a rather curious example:

“In ancient times, the Phoenicians cut [the cedars of Lebanon] from the mountains and thus brought devastating floods and silt down to the valleys below. Can we not say that the Phoenicians, by this policy, defaulted in their duties to the present inhabitants of Lebanon? […] It would seem that the duty to protect the right of the present Lebanese to have the cedars applied to those who were in a position to protect this right: the ancient Phoenicians” [16, p. 54].

The present Lebanese have no present rights to the trees, as there is no rational agent alive who is in a position to protect this right. For Partridge, a right exists if and only if some existing person is in a position to interfere with (the rightsholder’s) interests [16, p. 41]. Partridge’s account poses a lot of questions. How can nonexistent beings be “treated” in a certain way? Does this account not create infinite moral obligations many of which one is oblivious of? And does this account lack plausibility as it seems to give equal weight to present and future generation’s rights?

The first point is crucial. We can counteract future persons’ interests in various ways. When facing moral decisions, we must therefore consider the probability that we violate a place holder’s right by acting in a certain way. We act contrary to others’ interests if our action decreases the probability that their interests will be fulfilled. This relationship between one person’s action and another person’s (even contingent) in-
terest constitutes the right of the latter. Epistemic and ontological uncertainties regarding rightsholders might have been exceptional cases a few centuries or even decades ago. Today, they are everywhere: moral relationships towards people that are unknown, whose number and even existence often remain unclear. Consumer choices might be a good example. Whether online shopping, banking, or big industry – globalisation and development have changed the world we live in. One does not, therefore, have to appeal to scepticism about other persons’ existence when stating that uncertainties are common.

These considerations might provide us with hints about how to tackle the remaining two issues. Is the approach impossible to put in practice? After all, we might violate future person’s rights all the time without noticing it (every act has potential implications for other persons). While this might be true, it is clear that in many cases, we know exactly what we are doing. Climate change, diminishing resources, overpopulation, nuclear waste, unjust and unsustainable market systems are just some examples of issues that are clearly harmful to future generations and pose a threat to their very subsistence. In those cases, it would follow that we are obliged to care for future generations.

The third question deserves more attention. Is the view too demanding in that present and future generations are both granted present rights, and how should they be distinguished? In talking about future rightsholders, we have to estimate the probability that future persons will exist and will be harmed considerably by our actions. If this probability is high enough (presumably above a certain threshold) we have to acknowledge that these persons are rightsholders. However, it should be noted that many contemporary discussions concern actions whose relevance for future persons’ wellbeing is not under question.

In these cases, we should remind ourselves, as Henry Shue notes, that the present generation also has basic rights that must be respected [17]. Steigleder argues that decisions concerning future persons should be made from an internal perspective of the present generation, since our own rights are at stake in those decisions as well. They therefore precede over similar interests of future generations [18]. Discount rates can thus be avoided, while at the same time the plausibility of the rights-based account can be preserved. This is arguably even more important and relevant when future generations are conceived of as present rather than future rightsholders.
4 Conclusion

The upshot of this text is that the nonexistence argument should be rejected. I argued that in explaining obligations to future persons with their future rights one faces a challenge in linking future rights to present obligations, especially if one wants to uphold a strong notion of future peoples’ rights. Alternatively, we could treat future generations as present rightsholders. This becomes especially plausible when we take into account the fact that ethics nowadays has to deal with more and more uncertainties. For example, the view could be applied to the challenge of climate change that was mentioned as an introductory example. The future generations that will suffer from man-made climate change, on this account, are the value that the placeholder, conceived as a variable, will take in the future. This placeholder can be ascribed present rights claims. The approach presented in this paper thereby helps to theoretically reinforce our intuitions that we have to strengthen our efforts to prevent climate change, not only because of its effect on this and the next, but also distant generations.

Acknowledgements

I owe my gratitude to Professor Klaus Steigleder for his very insightful comments on earlier drafts of this paper. Also, I would like to thank an anonymous reviewer for KRITERION – Journal of Philosophy for providing helpful remarks and pressing me on important issues.

Notes

1 Gewirth writes that a “right is an individual’s interest that ought to be protected for his own sake and controlled by him. Since future agents will have such interests, they have rights” [9, p. 209]. Davidson makes the same point. He argues that we can affect the interests of future people, and therefore their rights: “I see no logical problem in the existence of a present duty to respect possible future rights of future generations” [4, p. 475]. Steigleder speaks of the “truism that people who do not yet exist do not have rights” and asserts that nonetheless “duties accrue to us today from the rights the people living in the future will have” [18, p. 8].

2 Gosseries gives three examples of non-correlational duties, two of which are based on an interest theory of rights [10]: First, if many people have an interest that is not significant enough to constitute a right on an individual level, these interests,
taken together, might still ground a corresponding obligation. One problem with this case is that it is not clear why a significant collective interest would not constitute a collective right, which Gosseries wants to avoid. Also, it remains unclear whether these dispersed significant interests constitute rights and obligations that are as strong as those inferred from individual significant interests. Even then, it would not be obvious that obligations can be inferred from non-significant individual interests. This claim would need further justification. The second of Gosseries’ cases for non-correlativity is demandingness for the obligation bearers. But should states be in the position to grant or withhold right statuses? Obviously, this might be problematic. Gosseries’ third case, powerlessness, says that will theorists could not grant rights to powerless people. So if we are supposed to have obligations towards them, this must be based on their interests, not their rights. But seen in this way, would this not lead to an inflation of obligations similar to the “inflation of rights” that will theorists, according to Gosseries, want to avoid [10, p. 449]?

3 Interestingly, on first sight, this case poses a threat to both will and interest theorists: of course, past generations cannot anymore claim their rights, so it seems that will theorists cannot endorse the notion of the rights of past generations. On an interest account, it could be argued that past generations once had significant interest and rights, but this fact does not provide us with an explanation as to why we can have obligations towards them now.

4 Here, I disagree with Beckerman/Pasek, who argue that “properties, such as being green or wealthy or having rights, can be predicated only of some subject that can exist” [1, p. 20].

5 Unless we are deep ecologists, we would agree that it is not possible to have obligations towards a stone, or a room-palm tree, but that we have obligations towards a child, and that there is a criterion to distinguish these cases.

6 Similar examples have been brought forward by Parfit [15, p. 356]) and Beckerman [1, pp. 17-18].

7 A similar case is presented by Partridge. See [16, p. 56].

8 This is what Elliot calls the non-concessional view: „there is no present bearer of the right but that, nevertheless, the right exists now and its present existence is contingent on the future existence of some person who will then be the bearer of the right“ [6, p. 161].

9 It is possible to explain within the future-rights-for-future-persons-view why we have obligations towards future selves. However, for the reasons mentioned above I believe that this renders a weaker notion of future selves’ rights than one might (intuitively) want.

10 It is hard to determine when a person’s existence begins or ends. The most consensual account probably is that human beings do not start to exist before they are conceived, and do not cease to exist until they are brain dead. But there might be good reasons for quite different definitions. These are crucial questions to many of the topics currently discussed in applied ethics (for example, concerning abortion, organ transplantation, or, in some cases, medically assisted suicide). Still, for the purpose of this paper, I will leave these questions open, as it is not decisive to my argument where exactly the line between existing and not existing is drawn. It is also necessary to distinguish among non-existent persons people who have existed in the past from future generations. (All of these persons are actual, in contrast to merely possible people).
11 The example is similar to Elliot’s missile case. He defends the non-concessional view suggesting an analogy to future generations by arguing that if we fire a missile on an island we violate contingent rights of the potential inhabitants [6, p. 161].

12 See [16, p. 57]: “My duty not to be negligent is a duty to anyone who might be injured, and if my duty is fulfilled […] the rights of unidentified multitudes will thereby be respected.”

13 Additional considerations might come into play here when weighing rights claims. Those might, for example, include issues such as demandingness, emergency, personal bonds and proximity.

Charlotte Unruh
University of Reading
c/o St Patrick’s Hall
Northcourt Avenue
Reading RG2 7HB, United Kingdom
<charlotte.unruh@ruhr-uni-bochum.de>
References


Abstract
When Schlick introduced his theory of „Konstatierungen“ in the Vienna Circle’s Protocol-Sentence Debate, the sole other member of the Circle who took part on his side was his former disciple Béla Juhos. However, Juhos’ own conception, developed in full extend only later, differs in important respects from the one of his teacher. Both conceptions are critically examined in the first part of this paper. The second part is devoted to an examination of the charge that such basic propositions are by all means useless because of incommunicability.

Keywords: Vienna Circle, basic propositions, intersubjectivity

Einleitung

http://www.kriterion-journal-of-philosophy.org © 2016 The author
Juhos blieb Schlick Zeit Lebens verbunden; dies geht schon allein daraus hervor, dass nicht weniger als drei späte Arbeiten ([15], [17], [18]) exklusiv der Darstellung und Würdigung des Lehrers gewidmet sind. Der Tenor dieser Schriften ist der gleiche: Schlick habe den Grundstein für eine echte Neuorientierung der Erkenntnistheorie gelegt, seine bleibende Bedeutung sei durch inzwischen erreichte Ergebnisse zu speziellen Fragen nicht geschmälert, ja in entscheidenden Fragen sei er über Wittgenstein und Carnap hinausgegangen. Wenn in diesen Texten gelegentlich die Perspektive des Autors erkennbar wird, so spricht das umso mehr dafür, dass Schlick den bestimmenden Ausgangspunkt für Juhos’ eigene Entwicklung darstellt.

In der anderen Richtung betrachtet, stellt sich die Beziehung allerdings weniger positiv dar. Zwar kann kein Zweifel daran bestehen, dass Schlick Juhos jedenfalls bis zu einem gewissen Grad gefördert hat, als Lieblingsschüler kann er aber nicht bezeichnet werden. Die einzige mir bekannte Stelle, an der Schlick sich dezidiert über Juhos äußert, findet sich in einem Brief an Carnap, der (als Herausgeber der Erkenntnis) Schlicks Meinung zu einem von Juhos eingereichten Manuskript hören wollte.\(^3\) Die Antwort ist überraschend deutlich; das Manuskript hat dieselben Fehler wie seine früheren Arbeiten: J. ist ein Eigenbrötler, der die Dinge in falscher Perspektive sieht, weil er auf allerlei Unterscheidungen herumreitet, die man wohl machen kann, die aber unwichtig sind, von ihm jedoch für das Wesentliche gehalten werden.\(^4\)


mehr vergönnt, eine zusammenfassende Darstellung seiner Erkenntnisto-
theorie rund um den Zentralbegriff der Konstatierungen zu verfassen. Ich werde hier, wo es vorrangig nicht um die Rekonstruktion der Ent-
wicklung dieser Konzeption geht, auf solche Divergenzen nicht eingehen,
sondern mich auf weitgehend unstrittige Eckpunkte beziehen.⁸

Beim Juhos dagegen liegt die Sache anders: In seiner 1950
veröffentlichten (1948 als Habilitationsschrift eingereichten) Monogra-
phie Die Erkenntnis und ihre Leistung [14] liefert er eine systematische
Darstellung seiner erkenntnistheoretischen Auffassungen; ausdrücklich
bezeichnet er eine ganze Reihe von älteren Arbeiten als Vorarbeiten zu
dieser Schrift.⁹ Wir haben also, was die Konstatierungen anlangt, ei-
eine umfassende systematische Darstellung vor uns; und soweit ich sehen
kann, modifiziert Juhos in dieser Hinsicht seinen Standpunkt auch im
weiteren Verlauf nicht mehr. Im Folgenden stütze ich mich, was Juhos
betrifft, also auf dieses Buch.

Einigkeit besteht zwischen Schlick und Juhos im Wesentlichen, was
die Rolle der Konstatierungen anlangt: Sie sind Fälle von irrtumssiche-
er Erkenntnis, stellen den Endpunkt im Verifikationsprozess dar, durch
ihren unmittelbaren Realitätsbezug verbinden sie Theorie und Erfah-
runung, und, nicht zu vergessen, durch die Konstatierungen erhalten alle
übrigen synthetischen Aussagen überhaupt erst Sinn. Kurz gesagt, die
Konstatierungen bilden für beide sowohl die epistemische als auch die
semantische Basis. Gemäß dieser phänomenalistischen Einstellung sind
sich beide einig in der Frontstellung gegen den physikalistischen Flügel
des Wiener Kreises, oder, wie Juhos sich verallgemeinernd ausdrückt, die
„Hypothetisten“, die einen durchgängigen Fallibilismus vertreten.

Die folgenden Ausführungen sind in vier Abschnitte gegliedert. Die
ersten beiden Abschnitte sind der epistemischen Sonderstellung der Kon-
statierungen gewidmet. Der Schwerpunkt liegt dabei auf Juhos, einfach
weil seine Ansichten viel weniger bekannt sind als die seines Lehrers.
Im dritten Abschnitt werde ich der Frage nachgehen, wie Schlick und
Juhos auf das wahrscheinlich wichtigste Argument der Gegenseite rea-
gieren, wonach Konstatierungen nicht intersubjektiv verständlich und
damit für die intersubjektive Wissenschaft jedenfalls ohne Belang sei-
en. In dieser Frage, so glaube ich, stellt Juhos’ Versuch der Entgeg
nung eine Ausarbeitung von Anregungen Schlicks dar, die sich verstreut in
publizierten wie unpublizierten Schriften der allerletzten Jahre finden.¹⁰
Diese Schlick/Juhos-Position ist aber m.E. ernsten Bedenken ausgesetzt.
Im letzten Abschnitt möchte ich eine Alternative zu diesem Vorschlag
zumindest skizzenhaft entwerfen, also versuchen zu zeigen, dass Konsta-
tierungen als Aussagen über das Vorliegen prinzipiell privater Erlebnisse durchaus intersubjektiv verständlich sind.

1 Die Irrtumssicherheit der Konstatierungen

Sowohl Schlick als auch Juhos sind sich darüber einig, dass die Irrtumssicherheit der Konstatierungen auf deren Grammatik beruht, also auf den Regeln, die für den Gebrauch von sprachlichen Zeichen gelten. Gemeinsamer Ausgangspunkt und gemeinsames Ziel verhindern allerdings nicht, dass die Argumentationen ganz verschieden laufen.


Irrtum kann es nur geben, wo es nachzuprüfende Prognosen gibt.

Aus Konstatierungen lassen sich überhaupt keine Prognosen ableiten.

Ergo: Bei Konstatierungen besteht keine Möglichkeit des Irrtums.

Ich glaube nicht, dass dieses Argument einer Analyse standhält. Ich werde mich auf drei Punkte beschränken:

Zum Ersten scheint mir hier wieder jene Vermengung von Wahrheitsbegriff und Wahrheitskriterium vorzuliegen, von der schon die Protokollsatzdebatte angekränkelt ist. Denn auch hier wird ja daraus, dass es kein Kriterium gibt, darauf geschlossen, dass keine Falschheit, also kein negativer Wahrheitswert vorliegt. Diese unzulässige Vermischung führt dazu, dass Juhos, scheinbar ohne es zu merken, eigentlich das Thema wechselt. Argumentationsziel ist ja ursprünglich, zu beweisen, dass ein Konstatierender notwendigerweise eine wahre Aussage macht. Was das Argument aber bestenfalls zeigt, ist, dass eine Konstatierung durch nachfolgende


Drittens scheint es mir überhaupt eine fehlgeleitete Strategie zu sein, die mangelnde Ableitbarkeit anderer Sätze aus Konstatierungen so ins Zentrum zu stellen, wie Juhos das tut. Zuerst einmal ist fest-

2 Sprechen und Glauben

Der letzte Punkt leitet über zu einem Folgeproblem, denn im Zusammenhang der Unterscheidung von sprachlichem und sachlichem Zweifel unterminiert Juhos selbst die These, wonach aus Konstatierungen nichts Weiteres zu folgern ist.\textsuperscript{18} Diese Unterscheidung ist offensichtlich notwen-


Wie nimmt sich dieses Verhältnis von Konstatierung und Erinnerungsglaube, von sprachlichem und sachlichem Irrtum bei Schlick aus? Bei ihm heißt es:


Jeder assertorische Satz bringt nicht nur den Glauben des Sprechers zum Ausdruck über das, worüber geurteilt wird, sondern drückt auch einen Glauben aus bezüglich der Wörter, die im Satz verwendet werden. Der Satz „Dies ist Rot“ drückt nicht nur meine Meinung bezüglich einer


Bevor wir jedoch weitergehen, ist es wohl angebracht, zumindest einen kurzen Blick auf die – zumindest was Schlick betrifft – doch recht intensiv geführte Diskussion zu werfen. Bereits eingangs erwähnt (und außer Streit stehend) ist die Tatsache, dass sich nicht alle Aussagen Schlicks zum Thema Konstatierungen in ein kohärentes Ganzes bringen lassen; dementsprechend divergierend fallen die verschiedenen Interpretationsansätze aus.21 Als ein besonders radikales Beispiel sei hier Uebel [47] erwähnt, der die epistemologische Funktion völlig herunterspielt und die Konstatierungen ausschließlich als semantisches Bindeglied zwischen Sprache und Welt interpretiert, in der Art von „unsagbaren Erläuterungen“ im Sinne von Wittgensteins Tractatus. Im Allgemeinen gilt „Über das Fundament der Erkenntnis“ ja als ein Schlüsseltext für den erkenntnistheoretischen Fundamentalismus im 20. Jahrhundert, und Uebels Interpretation hat dann auch jede Menge Textstellen gegen sich, in denen Schlick eindeutig von der Rolle der Konstatierungen im Prüfverfahren spricht (für eine entsprechende Kritik an Uebel siehe Oberdan [28]). Aber auch vom hier entwickelten Ansatz aus muss Uebels Interpretation zurückgewiesen werden: Als eigentliche Konstatierung wurde ja eben das unmittelbare, vorsprachliche Bemerken eines Erlebnisses herausgearbeitet – ob das bemerkte Erlebnis dann für eine hinweisende Definition eines Erlebniswortes verwendet wird oder nicht,
die Konstatierung selbst ist keine hinweisende Definition, keine bloße „Erläuterung“, sondern genuine Erkenntnis.


Schwieriger abzuwehren ist ein Einwand, den Lehrer und vor ihm (zumindest in nuce) bereits Scheffler erheben ([21, pp. 55–58], [33, pp. 111–114]). Die Irrtumssicherheit der Konstatierungen ist nach Schlick ja darin begründet, dass der Sinn von hinweisenden Ausdrücken nur verstanden werden kann, indem man sie gleichzeitig anwendet, verifiziert. Nun kann (jedenfalls um des Argumentes willen) zugegeben werden, dass der Akt einer hinweisenden Referenz zwar eine notwendige Bedingung für die Verifikation von Konstatierungen ist, aber, so Lehrer, es handelt sich um keine hinreichende Bedingung. Eine Konstatierung enthält ja neben deiktischen Ausdrücken auch deskriptive, und es fehlt noch jede Garantie dafür, dass das Erlebnis, auf das hinweisend referiert wird, tatsächlich unter den verwendeten deskriptiven Begriff fällt. Damit zielt
dieser Einwand direkt gegen die Gültigkeit der Schlick’schen Formel, womit im Falle von Konstatierungen aus dem Verstehen bereits Einsicht in die Wahrheit, d.h. Verifikation folgt. Tatsächlich scheint mir dieser Einwand schwer abzuweisen zu sein. Dass im unmittelbaren Bemerken bereits Begriffsverständnis enthalten ist, bleibt davon jedenfalls unberührt (bestritten wird nur, dass dieses Bemerken notwendigerweise ein wahres Urteil ist), und implizit gibt Lehrer auch zu, dass die Konstatierungen gegenüber allen anderen synthetischen Urteilen epistemologisch privilegiert sind: Sein Argument beruht darauf, dass Konstatierungen nicht nur hinweisend referieren, sondern auch eine deskriptive Komponente haben. Die Unmöglichkeit einer Fehlreferenz wird auch bei ihm nicht bestritten, so dass zumindest in dieser Hinsicht ein Privileg besteht. Konstatierungen können nicht schon deshalb falsch sein, weil dasjenige, worüber geurteilt wird, schlichtweg nicht existiert (wie im Falle der Außenwahrnehmung bei Totalhalluzination jederzeit möglich).


Noch ganz kurz zum zweiten, intensiv diskutierten Problemkomplex, der die Rolle der Konstatierungen im Prüfprozess betrifft. Die wesentliche Schwierigkeit dabei ist zu verstehen, wie die außerhalb des wissenschaftlichen Satzsystems stehenden Konstatierungen mit diesem zusammenhängen. Um eine Rolle im Prüfprozess spielen zu können, müssen doch logische Beziehungen zwischen nachzuprüfenden Hypothesen und
Konstatierungen bestehen, so scheint es jedenfalls. Unmittelbar nach Lektüre von „Über das Fundament der Erkenntnis“ schreibt Carnap an Schlick:

Besonders ist mir unklar, wie die Ableitung solcher Sätze [der Konstatierungen, J. F.], in denen „hier“ und „jetzt“ vorkommt, und zu denen wesentlich hinweisende Gebärden gehören, aus einer nachzuprüfenden Hypothese aussehen soll; diese Ableitung ist ja für die Nachprüfung sicherlich erforderlich.  


3 Die intersubjektive Verständlichkeit der Konstatierungen

Es sollte klar sein, dass jede Verteidigung der Irrtumssicherheit der Konstatierungen die Vorbedingung hat, dass solche Urteile keinerlei Bezugnahme auf Physisches implizieren dürfen. Konstatierungen handeln nur von Erlebnissen und besagen nichts über irgendwelches Physisches. Aufgrund dieser unabdingbaren Voraussetzung konstruieren die Opponenten in der Protokollsatz-Debatte ihr bekanntes Grundsatzargument, das sich nicht nur gegen die Konstatierungen, sondern gegen alle Aussagen über rein Psychisches richtet:


Es sieht so aus, als habe Schlick anfänglich diese Situation noch für unausweichlich, aber auch für akzeptierbar gehalten. So heißt es etwa, dass Konstatierungen nicht aufgeschrieben (und das heißt wohl: nicht mitgeteilt) werden können [35, p. 513]. Aber das ist eine unhaltbare Position, denn die Konstatierungen sollen ja nicht nur den epistemischen Endpunkt der Verifikation darstellen, sie bilden ja auch die semantische Basis für alle übrigen empirischen Aussagen (Hypothesen): Der Sinn von Hypothesen wird durch ihre Verifikationsmethode bestimmt, die letztlich zu Konstatierungen führt. Unabhängig von der exakten Formulierung sagt das Verifikationsprinzip im Kern, dass jede synthetische Aussage nur durch die Anbindung an die Erfahrung Sinn hat; vorsichtig ausgedrückt
heißt das, dass alle inhaltlichen Begriffe aus der Erfahrung stammen und entsprechende Ausdrücke nur durch Hinweis auf Erfahrung in die Sprache eingeführt werden können – wer selbst keine Farb-Erlebnisse hat, dem kann keine Definition oder Erörterung die Bedeutung von „rot“ klarmachen, wie Schlick es bereits in der *Allgemeinen Erkenntnislehre* in aller Deutlichkeit herausstellt [42, p. 199f.]. Ohne diese semantische Basis sind Hypothesen bloße Leerformeln, man kann gar nicht sagen, wonach sie eigentlich handeln.


Verteidiger der Konstatierungen stehen damit vor der Aufgabe, zu zeigen, wie Konstatierungen intersubjektiv verständlich sein können. Die Antwort von Schlick und Juhos, der sich seinem Lehrer in diesem Punkt anschließt, besteht im Verzicht auf Privatheit, genaugenommen in der These, dass es sich bei der Privatheit von Sinnesdaten nicht um eine prinzipielle, sondern nur um eine empirische Tatsache handelt. Ich gebe die Argumentation in Kurzform wieder:


Anstatt solche und weitere Argumente in gebührender Präzision und Ausführlichkeit zu entwickeln, möchte ich im abschließenden Teil versuchen, eine Alternative zumindest skizzenhaft zu umreißen. Also versuchen zu zeigen, dass Konstatierungen intersubjektiv verständlich sind,

4 Intersubjektivität: Skizze einer Alternative

Es gibt einen Punkt, in dem sich Physikalisten und Schlick/Juhos einig zu sein scheinen: Für die intersubjekive Verständlichkeit einer Tatsachenaussage sei es notwendig, dass sich verschiedene Personen auf dieselbe Weise auf denselben Sachverhalt beziehen können. Die verschiedenen Lösungsvorschläge wollen beide dieser Voraussetzung gerecht werden: Der Schlick/Juhos-Ansatz durch Ablehnung einer prinzipiellen Privatheit und damit der These, dass sich (zumindest der Denkmöglichkeit nach) verschiedene Subjekte auf dieselbe Weise, nämlich konstatierend, auf das Vorliegen desselben Sinnesdatums beziehen können. Laut physikalischem Vorschlag werden Erlebnis-Aussagen umgedeutet in Aussagen über Physisches; und auf physische Sachverhalte können sich verschiedene Personen in gleicher Weise beziehen.


Schlick/Juhos vertreten also, dass eine Konstatierung von verschiedenen Subjekten zur Referenz auf denselben Sachverhalt verwendet wer-
den kann. Im Fall der Konstatierungen heißt das, dass „Ich habe eine Rotempfindung“, geäußert von A und B, nicht notwendigerweise bedeutungsverschieden ist. Bedeutungsgleichheit besteht dann, wenn das Sinnesdatum, auf das sich A und B im gleichen Referenz-Modus beziehen, dasselbe ist. Da dies zumindest der Fall sein kann, soll sich ergeben, dass die Voraussetzung erfüllt ist: A und B können sich mit bedeutungsgleichen Sätzen auf dasselbe beziehen, und so ergibt sich die intersubjektive Verständlichkeit der Konstatierungen.


Die Voraussetzung, dass für intersubjektive Verständlichkeit der Konstatierungen verschiedene Subjekte bedeutungsgleiche Sätze äußern können, kann nicht erfüllt werden, sie braucht es aber auch gar nicht. Um die Konstatierung einer anderen Person verstehen zu können, muss der Verstehende aus seinem eigenen Fall wissen, was es heißt, sich selbst direkt hinweisend ein Sinnesdatum zuzuweisen. Versteht man die semantische Funktion von „ich“ – die Eigentümlichkeit der Referenz im Ich-Modus – so versteht man auch Sätze anderer Personen mit diesem Pronomen. Wir haben dann als sozusagen objektiven Gegenpart einen Satz wie etwa „die so und so identifizierte Person schreibt sich selbst in direkt hinweisender Form zu, eine Rot-Empfindung zu haben“. Offensichtlich ist dieser Satz nicht bedeutungsgleich mit der Fremd-Konstatierung, aber das Verständnis scheint damit adäquat zum Ausdruck gebracht. Obwohl der andere meine Konstatierung nicht machen kann – und genau das wird eigentlich von Schlick/Juhos gefordert – kann er sie verstehen. Die These der bloß empirischen Privatheit von Erlebnissen (samt der „Körper-Theorie des Ich“) ist schon wegen der in
Abschnitt 3 herausgestellten Einwände abzulehnen, sie kann aber auch nicht leisten, was sie laut Schlick/Juhos leisten soll. Das ist aber, wie nun zumindest angedeutet, auch gar nicht notwendig: Die Konstatierung einer anderen Person ist für mich verständlich, obwohl das entsprechende Erlebnis ein Erlebnis nur dieser Person ist.

Eine genauere Explikation der semantischen Besonderheiten der Ich-Referenz kann hier nicht erfolgen, es sei an dieser Stelle nur auf die mittlerweile weithin akzeptierte Tatsache hingewiesen, dass im sogenannten emphatischen Selbstbezug „ich“ nicht durch andere, „objektive“ Termi

35


Die Argumentation bezog sich bisher nur auf den Referenz-Modus. Das ist aber nur die eine Seite der Medaille, es geht ja nicht nur um die Art der Bezugnahme, sondern auch darum, was auf verschiedene Weise zugeschrieben wird. Diese zweite Seite betrifft also das Verständnis des zugeschriebenen Prädikats, in diesem Beispiel „rot“. Während hier für die notwendige Verschiedenheit des Referenz-Modus argumentiert wurde (woraus aber keine Nicht-Mitteilbarkeit folgt), darf kein entsprechen der Unterschied hinsichtlich der zugeschriebenen Eigenschaft bestehen. Der Asymmetrie der Referenz darf keine Asymmetrie der zugeschriebenen Prädikate entsprechen. Wäre die Bedeutung von „rot“ verschieden je nachdem, ob eine konstatierende Person sich selbst diese Eigenschaft zuschreibt oder jemand anderem, so wäre tatsächlich keine Verständigung mehr möglich. „Rot“ darf also nicht bei Selbstzuschreibung für eine rein psychische Erlebnisqualität, bei Fremdzuschreibung für physisches Verhalten stehen, ansonsten würden verschiedene Subjekte systematisch a ineinander vorbeisprechen. Da nun die physikalistische Symmetrie (in allen Fällen ist nur Verhalten gemeint) m.E. völlig unhaltbar ist, muss die Lösung in der anderen Richtung gesucht werden. Ich kann hier al-
lerdings nur mehr ganz dogmatisch behaupten, dass mir diesbezüglich die alte Methode des Analogieschlusses als der richtige Weg erscheint: „Rot“ steht immer für eine Erlebnisqualität, und wir nehmen an, dass im Fall einer Fremdzuschreibung ein qualitativ ähnliches Erlebnis (ein Erlebnis der gleichen Art) vorliegt wie dasjenige, das wir bei Selbstzuschreibung unmittelbar bemerken und von dem wir damit unmittelbares Begriffverständnis haben.

Diese Skizze (und um mehr handelt es sich nicht) einer Lösung des Problems der intersubjektiven Verständlichkeit der Konstatierungen stellt im Grunde nur eine Verteidigung unserer Common-Sense-Auffassung dar. Wir glauben nicht, dass eine Aussage über ein Erlebnis nichts anderes ist als eine Aussage über körperliches Verhalten, so wie der Physikalismus das tut. Ebenso wenig glauben wir aber, dass es für das Verständnis von Erlebnis-Aussagen anderer erforderlich ist, sich direkt hinweisend auf dieses fremde Erlebnis beziehen zu können, so wie es der Schlick/Juhos-Ansatz postuliert. Diese beiden konträren Ansätze widersprechen nicht nur dem Common Sense (was für sich genommen noch kein Argument ist), sie halten auch beide einer Prüfung nicht stand. In diesem Sinn haben die Ausführungen zum Intersubjektivitätsproblem nur das Ziel, dazu beizutragen, dass der aufgewirbelte Staub sich wieder legt.40

Notes


2 Eine vollständige Bibliographie ist enthalten in [19].

3 Rudolf Carnap an Moritz Schlick, 29. Juni 1934, Nachlass Schlick; das Manuskript ist jedenfalls nie unter dem ursprünglich vorgesehenen Titel („Der Zufall als praktische und methodische Voraussetzung der idiographischen Wissenschaften“) erschienen, vermutlich finden sich dort ausgearbeitete Gedanken aber in späteren Texten wie etwa [13].


5 Die Verleihung des Titels eines außerordentlichen Professors 1955 stellte die letzte Stufe in Juhos’ akademischer Karriere dar. Wie Juhos selbst nicht ohne Bitterkeit bemerkte, wurden „die Vertreter der erkenntnisanalytischen Philosophie […]“
Johannes Friedl: Erkenntnisbasis und Intersubjektivität

nur als unbesoldete Dozenten zugelassen und sehen sich so, wenn sie durch ihre Lehrtätigkeit den Lebensunterhalt verdienen wollen, zur Abwanderung genötigt“ [17, p. 132]. Nur die wirtschaftliche Unabhängigkeit Juhos’ ersparte ihm dieses Schicksal; vgl. dazu wie überhaupt zur Biographie [29].

6 Freilich hatte auch Juhos Schüler und blieb insofern nicht ohne Wirkung. Neben Hubert Schleichert, dem Verfasser des angeführten Nachrufs [34], ist hier mindestens noch Werner Leinfellner zu erwähnen; weitere Schüler sind bei Reiter angeführt [29, p. 85]. Dort (p. 86) findet sich auch die für Juhos’ Position aufschlussreiche Information, dass Schleichert auf die Mitautorenschaft für ein gemeinsam mit Juhos erarbeitetes Werk [16] verzichtete, um nicht seine Chancen auf eine Habilitation in Wien zu schmälern.


8 Für eine Rekonstruktion der Entwicklung und der (bei Schlick nicht gänzlich zufriedenstellend aufgelösten) Spannungen siehe [7, Kap. V].


10 Siehe dazu unten, Anm. 33.

11 Im Folgenden werde ich mich an die erstgenannte Formulierung halten; die (sehr unglückliche) „verblose“ Formulierung „Hier jetzt rot“ [35, p. 511ff.] wird auch von Schlick selbst im weiteren Verlauf nicht mehr verwendet.

12 Zwar stellt auch Koterski diesen Unterschied fest [20, p. 166], aber er arbeitet die Differenzen zwischen Schlick und Juhos m.E. unzureichend heraus.

13 Alston unterscheidet neben Irrtumssicherheit zwischen Unbezweifelbarkeit (es gibt keine konträre Evidenz) und Unkorrigierbarkeit (die Evidenz, die für eine Konstatierung spricht, ist jedenfalls stärker als jede konträre Evidenz) [1, insbesondere p. 384].


15 „Logisch schließt sich nichts mehr an sie an, es werden keine Schlüsse aus ihnen gezogen, sie sind ein absolutes Ende.“ [35, p. 507].

16 Ein hier nicht weiter zu erörterndes Problem besteht in der Vollständigkeit und der genauen Exklosion solcher „Normalbedingungen“ der Wahrnehmung.

17 Dagegen macht es durchaus Sinn, zu fragen, woher man weiß, dass ein Satz tatsächlich irrtumssicher ist. Und wird auf diese Frage mit der Angabe irgendeines Merkmals X geantwortet, das einen solchen Satz auszeichnet, kann diese

18 Zum Folgenden vgl. [14, p. 139f.].


20 Und nicht etwa Sätze, die Erinnerungen ausdrücken: Damit würde sich ja sofort wieder das Problem der richtigen Sprachverwendung stellen und ein infinite Regress in Gang gesetzt.

21 Für eine Übersicht über die verschiedenen Interpretationen siehe [48, pp. 361–370].


24 Närher ausgeführt ist das Folgende in [7, Kap. V, Ab. 4].

25 Für Näheres dazu siehe wiederum [7, p. 234ff.]

26 Am klarsten bei Carnap [3, Abschnitt 6].

27 Für diese Problemexposition vgl. Rutte [31, p. 179f.]; diesem Aufsatz verdanken auch die folgenden Ausführungen maßgebliche Anregungen.

28 Die Forderung der Definierbarkeit (im strengen Sinn) aller inhaltlichen Ausdrücke durch einfache Erlebnisbegriffe dagegen – das kann als Resultat der bekannten Diskussion in der weiteren Entwicklung des logischen Empirismus festgehalten werden – muss als unerfüllbar fallengelassen werden.


31 Die Argumentation bis hierher stammt ursprünglich von Wittgenstein, was von Schlick an genannter Stelle auch explizit gemacht wird.

32 Juhos [14, p. 248]; offensichtlich ist hier numerische Identität und nicht bloß qualitative Gleichheit gemeint.

33 An dieser Stelle scheint es angebracht, die verschiedenen ineinandergreifenden Teile von Schlichs letzten Standpunkt zum Intersubjektivitätsproblem zu lokalisieren: Neben [36], wo sich die eben dargestellte „Eliminierung“ des Subjekts findet, ist [37] von einschlägiger Relevanz, als der Ort, an dem Schlick seine Kritik am semantischen Physikalismus formuliert. Und auch der Analogieschluss auf
Fremdpsychisches erfährt eine Rehabilitierung, da nun der alte Einwand, wonach dieser Schluss wegen der Unüberprüfbarkeit der Konklusion nicht sinnvoll sei, abgewiesen werden kann; zum letzten Punkt vgl. eine bislang unpublizierte Vorlesung Schlicks [41, p. 65f.]. Zur Stützung dieses interpretativ hergestellten Zusammenhanges vgl. auch die „der Darlegung der Lehrmeinungen Prof. Schlicks“ gewidmete Arbeit seiner Schüler Heinrich Melzer und Josef Schächter [24, Zitat p. 92].

34 Und damit etwas zu versuchen, was andere bereits besser getan haben, als ich es vermag; vgl. insbesondere Rutte [32]. Vgl. auch das Argument von Ambrus [2, p. 126f.], das eine innere Widersprüchlichkeit dieses Standpunktes herauszuarbeiten versucht.

35 Näheres dazu in Rutte [32].

36 Vgl. dazu die zusammenfassende Darstellung von Frank [6, Abschnitt IV].

37 Siehe das Zitat oben, S. 8; vgl. dazu auch Chisholm [4].

38 An dieser Stelle zumindest zu erwähnen ist der Funktionalismus, der versucht, die Dichotomie durch eine „neutrale“ Analyse von (im üblichen Sinn verstandenen) psychologischen Begriffen zu unterlaufen.

39 Wie bereits kurz erwähnt (siehe Anm. 33), scheint auch Schlick letztlich den Analogieschluss wieder zu akzeptieren, allerdings nur unter der (hier zurückgewiesenen) Voraussetzung, dass die Konklusion eines solchen Schlusses zumindest prinzipiell direkt verifizierbar ist.

40 Ich danke Heiner Rutte für die Diskussion früherer Versionen dieses Textes und die daraus resultierenden zahlreichen Verbesserungsvorschläge; ebenfalls zu Dank verpflichtet bin ich einem/r anonymen Gutachter/in.

Johannes Friedl
Universität Graz
Institut für Philosophie
Heinrichstraße 26
8010 Graz, Austria
<johannes.friedl@uni-graz.at>
Literatur


[38] Moritz Schlick. „Introduction“. In: [44], pp. 395–399.


This year the Salzburg Conference for Young Analytic Philosophy (SOPhiA, Salzburger Concilium Omnibus Philosophis Analyticos) took place from September 7 to 9. As the title of the conference indicates, it is mainly organized by and aimed at graduate students in philosophy. Since 2010 the conference has been held annually in Salzburg and features contributions from many international participants. One of the distinctive features of SOPhiA is that it is open to contributions from every philosophical discipline—the only requirement being that the presented work is carried out in an analytic way. Analytic Philosophy is here broadly understood in the way expressed by Carnap in *The Logical Structure of the World* (1967): analytic philosophers ”have taken the strict and responsible orientation of the scientific investigator as their guideline” and address specific problems.

The conference featured more than 100 contributed papers that (broadly speaking) covered topics from all philosophical subdisciplines. The main sections were Epistemology, Metaphysics and Ontology, Ethics (approx. 20% each), Philosophy of Mind, Philosophy of Language, and Philosophy of Science (approx. 10% each). The remaining sections covered topics from Logic and Philosophy of Mathematics, Political Philosophy, History of Philosophy, and Philosophy of Law. The papers were given by graduate students (approx. 30% female) from 19 different countries. Some of these contributions were handed in as full papers of which a selection is published in this journal. The best paper award went to Maximilian Kiener (Oxford) for his paper on moral and non-moral testimony.

There were four plenary lectures given by invited speakers: Julien Murzi (Salzburg), Jeremy Butterfield (Cambridge), Dorothy Edgington (Birkbeck College), and Sylvia Wenmackers (KU Leuven). In the following the lectures shall be briefly summarized.

The conference started with a talk by Julien Murzi who challenged revisionary approaches to semantic paradox. Consider for example the Liar Paradox. In order to preserve consistency (or at least non-triviality) when dealing with the Liar Paradox, there are two options: (1.) either
give up on naïve semantic principles such as the equivalence of ’A’ and ’A is true’, or (2.) revise the rules of classical logic. Revisionary approaches choose option (2.), for instance, in the case of paracomplete approaches which allow for sentences having an intermediate value between truth and falsity. Murzi argued that all known revisionary approaches run into problems when dealing with semantic paradox. In particular, he argued against moderate revisionary approaches. These approaches attempt to restrict non-classical logic to the contexts of paradoxicality while trying to recapture classical logic for other contexts (such as mathematics). According to Murzi, these approaches rely on classifying contexts as paradox or unparadox. Murzi argued that this gives rise to a revenge phenomenon which is common to all revisionary approaches: (a) consistent non-classical theories fail to express a suitable notion of paradoxicality, and (b) inconsistent (but non-trivial) non-classical theories cannot express unparadoxicality. Thus, moderate revisionary approaches fail in their attempt to demarcate classical from non-classical contexts.

On the second day Jeremy Butterfield gave a talk on the implications of modern cosmology to scientific realism. In Butterfield’s approach scientific realism is to be understood as the claim that we can know, and indeed do know, about the unobservable. Butterfield takes cosmology as a particularly promising case study for an evaluation of scientific realism because it challenges the usual philosophical distinction between two types of under-determination: (1.) under-determination by all data that one could in principle obtain and (2.) under-determination by all data obtainable in practice. In his presentation Butterfield focused on primordial cosmology which investigates very early states of the universe (less than $10^{-11}$ seconds after the Big Bang). In particular he pointed out the stupendous success of cosmology in recent decades. Nevertheless, he noted, it is unsurprising that probing the very early universe involves intractable cases of under-determination of theory by data. This is particularly the case with regard to the inflationary period of the universe. However, when scientific realism is understood in the modest sense that we can know something (rather than everything) about the unobservable, then it is not threatened by modern cosmology. According to Butterfield, cosmologists make claims about the history of the universe which are and will forever remain as well established as other scientific facts, for instance, from molecular biology.

The second day of the conference ended with a talk by Dorothy Edgington on conditionals, uncertainty and indeterminacy. She began her talk by pointing out many situations in which we are confronted with
the vagueness of language, for instance, when we are presented with a colour spectrum and asked to indicate where one colour ends and the next begins. Frege, Russell and the early Wittgenstein aimed at constructing an ideal language that works without such borderline cases. However, according to Edgington this was a futile attempt: we cannot eliminate vagueness from natural languages and it is not even desirable. But how can we account for vagueness in a theory of language? Edgington suggested an idealization which assigns a number between 1 and 0 for the degree to which a statement is clearly true or clearly false. Moreover, she suggested giving these values a degree-theoretic structure that is analogous to probability theory. This analogy is to be taken with caution: probabilistic indeterminacy is often dynamic (the indeterminacy collapses, for instance, when we toss a coin) whereas indeterminacy through vagueness is static. Edgington then argued that there is another case of static indeterminacy which is very similar: the indeterminacy of counterfactual probability. According to Edgington, this analogy can be used for a better understanding of uncertainty in conditionals. Uncertain conditional judgements can be assessed by conditional probability. In this framework their truth value is often indeterminate, just as in the case of vague statements.

The conference concluded with a plenary talk by Sylvia Wenmackers on infinitesimal probabilities. Standard probability theory (Kolmogorov’s axioms) entails the axiom of continuity. As a consequence, probability values can be represented by real numbers which makes probability theory easily fit in with real analysis and measure theory. Wenmackers pointed out that the axiom of continuity is in conflict with the intuition that we should assign strictly positive probability to any possible outcome, no matter how improbable it is. This is problematic for example in the context of infinite lotteries (such as the lottery on \( \mathbb{N} \)) because the probability of any particular ticket winning has to be set to zero. The usual reaction to this conflict is dismissing the intuition and rounding tiny probabilities off to zero in order to preserve mathematical convenience. However, according to Wenmackers this is too high a price to pay. She argued that probability theory has to take remote contingencies into account and that without such infinitesimals probabilities do not add up. According to Wenmackers, these issues motivate adopting a non-Archimedean probability theory. Such a set of axioms gives up on the axiom of continuity and allows us to assign non-zero infinitesimal probabilities to remote contingencies. Wenmackers argued that this alternative approach to probabilities is unobjectionable from a
mathematical point of view but has attracted philosophers’ criticism. In particular, Wenmackers took up conceptual objections by Williamson, Easwaran and Pruss and argued that they do not threaten her non-Archimedean approach.

The conference was complemented by two affiliated workshops that took place on September 7: ”Dispositions in Action: Laws of Nature, Explanation and Modality” and ”Equivalence and Reduction of Scientific Theories”. Each workshop consisted of five talks by invited speakers. The workshops were kindly supported by the Gesellschaft für Wissenschaftsphilosophie (GWP).

SOPhiA 2016 was sponsored by the Gesellschaft für analytische Philosophie (GAP), mentis, KRITERION – Journal of Philosophy, Springer, Metzler, Suhrkamp, the University of Salzburg, the STV Philosophie of the University of Salzburg, ÖH Salzburg, Land Salzburg and Stadt Salzburg. The financial support made it possible for the conference registration to be free of charge and covered a warm buffet on the first evening. The conference was organized by Albert J. J. Anglberger, Simone Badergruber, Christian J. Feldbacher-Escamilla, Alexander Gebhart, Markus Hierl, Laurenz Hudetz, Sebastian Krempelmeier, Pascale Lötscher, Stefanie Orter, Mandy Stake, and Tobias Wagner. Thanks to the generous support, the organizing committee, the plenary speakers, workshop speakers and the many graduate student contributors SOPhiA 2016 was such a successful event.

Further information about SOPhiA 2016 is available at the conference website https://www.sbg.ac.at/sophia/SOPhiA/2016/languages/en/.

Enno Fischer
DFG-GRK 2073 ”Integrating Ethics and Epistemology of Scientific Research”
Institute of Philosophy
Leibniz Universität Hannover
Am Klagesmarkt 14–17
30159 Hannover
<enno.fischer@philos.uni-hannover.de>