EUROPEAN JOURNAL OF
ANALYTIC PHILOSOPHY

UDK 101 | ISSN 1845-8475

Publisher, Editorial office

University of Rijeka, Faculty of Humanities and Social Sciences, Department of Philosophy
Address: Sveučilišna avenija 4, 51000 Rijeka, Croatia
Phone: +385 51 265 794
Fax: +385 51 265 799
E-mail: eujap@ffri.hr

Web address

https://www.ffri.hr/phil/casopis/index.html

Printed by Impress, Opatija (Croatia)
200 copies available

CONTENTS

ARTICLES

FISSION, FIRST PERSON THOUGHT, AND SUBJECT-BODY DUALISM
Kirk Ludwig……..5

JUSTIFYING ONESELF
Mark Piper……27

A DEFENCE OF A RATIONALIST CONCEPTION OF PRACTICAL REASON
Gal Yehezkel…….39

"YES, THE THEORY IS ABSTEMIOUS, BUT . . .": A CRITIQUE OF YEHEZKEL
Regan Lance Reitsma ……..59

ABSTRACTS (SAŽECI) ……..80

INSTRUCTIONS FOR AUTHORS ……85

EuJAP | Vol. 13, No. 1, 2017
UDK
1:159.923.2
141.112
164.031

FISSION, FIRST PERSON THOUGHT, AND SUBJECT-BODY DUALISM*

KIRK LUDWIG
Indiana University

ABSTRACT

In "The Argument for Subject Body Dualism from Transtemporal Identity Defended" (2013), Martine Nida-Rümelin (NR) responded to my (Ludwig 2013) criticism of her (2010) argument for subject-body dualism. The crucial premise of her (2010) argument was that there is a factual difference between the claims that in a fission case the original person is identical with one of the successors. I argued that, on the three most plausible interpretations of 'factual difference', the argument fails. NR responds that I missed the intended, fourth interpretation, and that the argument on the third interpretation goes through with an additional assumption. I argue that the fourth interpretation, while insufficient as stated, reveals an assumption that provides an argument independently of considerations involving fission cases: in first person thought about future properties we have a positive conception of the self that rules out having empirical criteria of transtemporal identity. However, I argue that the considerations offered for this thesis fail to establish it, and that we do not, in fact, bring ourselves under any positive conception in first person thought, but rather think about ourselves directly and without conceptual mediation. This explains why it appears open in fission cases that the original person is identical with one of the successors, while what is possible is constrained by the actual nature of the self as referred to in first person thought. I argue also, incidentally, that on the third interpretation, the first premise of the argument is inconsistent with the necessity of identity.

Keywords: *subject-body dualism, fission cases, first-person thought, transtemporal identity, Martine Nida-Rümelin*

Acknowledgments:
My thanks to helpful conversation with Martine Nida-Rümelin and comments from two referees for this journal.

*Received: 25.6.2015.
Accepted: 17.8.2016.

1.Introduction

Imagine perfect fissioning of a person A into two successors who are equally good candidates for being the person who underwent the fissioning, LA (on the left) and RA (on the right), on the basis of all the material and psychological relations (empirical relations for short) that each bears to A. Imagine that if LA had been produced by the process without RA, we would judge that A = LA, and vice versa. We seem, *prima facie*, to understand what it would be for A to survive as (be identical with) LA (and not RA), and what it would be for A to survive as (be identical with) RA (and not LA). Does this show that

> [SBD] A is not identical with, or constituted by, or composed (even in part) by a body, or any material stuff, or even any immaterial stuff.

Call [SBD] the thesis of subject-body dualism.

In (Nida-Rümelin 2006) and (Nida-Rümelin 2010), Martine Nida-Rümelin (NR) argued that these considerations do establish [SBD]. The master argument has three premises:

1. There is a *factual difference* between the claim that someone is one or the other of the two continuers in fission cases or we are subject to a pervasive illusion in our thoughts about personal identity over time (call this *the illusion theory*).
2. There could be a factual difference between the claims that someone is one or the other of the two continuers in fission cases only if subject-body dualism were true.
3. The illusion theory is untenable (false).
4. Subject-body dualism is true [1-3].

Granting premise 3 for the sake of argument, we get the streamlined argument (when I refer to premises, I will have the streamlined argument in mind):

1. There is a factual difference between the claim that someone is one or the other of the two continuers in fission cases.
2. There could be a factual difference between the claims that someone is one or the other of the two continuers in fission cases only if subject-body dualism were true.
3. Subject-body dualism is true [1-2].

The basic idea of the argument is that the materialist can't make sense of there being a factual difference between the claim that someone is one or the other of two continuers in fission cases. The reason is that the materialist has to appeal to bodily or psychological continuity or some combination of these to ground claims about transtemporal identity (call

these empirical criteria of transtemporal identity), but in the fission case there is complete symmetry with respect to all the empirical criteria between A and RA, on the one hand, and A and LA, on the other. Whatever you can say about either case you can say equally about the other. How then can there be a factual difference between the two claims?

In (Ludwig 2013), I argued that on the three most plausible interpretations of 'factual difference' in premise 2, the argument failed. Consider our initial case, and let 'D' describe the relevant facts completely except for the facts about identity between A, LA, and RA. Then consider the claims P1-P3.

P1: D and A = LA.

P2: D and A = RA.

P3: D and A ≠ LA and A ≠ RA.

We assume that LA ≠ RA. In terms of this case, the three interpretations of 'there is a factual difference between P1 and P2' are these:

(1) P1 is true and P2 false *or* P1 is false and P2 is true.

(2) P1 and P2 differ in content (express different propositions).

(3) P1 and P2 both express genuine possibilities.

(1) says that there being a factual difference is there being a fact of the matter which is true and which false. (2) says that the factual difference is a matter of P1 and P2 expressing different propositions. (3) says that it is a matter of their expressing genuine possibilities (they might express different propositions but neither express something that is possible). Briefly, I argued in (Ludwig 2013) that (1) cannot be the right interpretation because NR allows that P3 is a genuine possibility and she doesn't argue it is not actual. I argued that (2) isn't sufficient for the argument because the materialist can make sense of there being a difference in content between P1 and P2 by saying that that 'A = LA/RA' means 'LA/RA's body is A's body'. Finally, I argued (in part—I return to this below) that (3) isn't sufficient because, even granting that, in a world in which P1 or P2, A would not have empirical criteria of transtemporal identity, it would be a modal fallacy to infer we don't *actually* have empirical criteria of transtemporal identity.

NR responded in (Nida-Rümelin 2013) that (i) the criticism fails because it overlooks the intended interpretation of factual difference (let's call this the fourth interpretation, which I will explain below), and that (ii), in any case, the third interpretation is sufficient for the argument to go through, when supplemented with the assumption that objects have their transtemporal identity conditions essentially. I respond in this paper.

In section 2, I take up the fourth interpretation, identify the underlying assumption that motivates it (that in first person thought we have a positive conception of the self that excludes empirical criteria of transtemporal identity), and formulate an argument for subject-body dualism that rests directly on that assumption. In section 3, I argue that we do not have good reason to accept the assumption because relevant features of first person thought are accommodated by our referring to ourselves directly and not under any mode of presentation (or by description). In section 4, I respond to the claim that the argument on the third interpretation is successful when combined with the assumption that we have our criteria for transtemporal identity essentially. I argue that the third interpretation requires a contradiction be true because it is inconsistent with the necessity of identity, so that the only possibilities we could be thinking of are epistemic possibilities. I explain this in terms of the account of first person thought developed in section 3. Section 5 is a short conclusion.

2. The Fourth Interpretation and the Reformulated Argument

The fourth interpretation is expressed in the following passage:

> [4I] … the fourth (and originally intended) interpretation of the claim that there is a factual difference [between] P1 and P2 … can be made explicit in the following way: *there is an objective feature which would have to be realized* ***in addition*** *to what description D states for P1 to be realized, and there is (a different) objective feature which would have to be realized* ***in addition*** *to what description D states for P2 to be realized.* (Nida-Rümelin 2013, 705)

NR says that (i) "that there is a factual difference between P1 and P2 is compatible with the metaphysical *impossibility* of both P1 and P2," for (ii) it "does *not* imply that the fulfillment of description D is metaphysically compatible with the relevant feature which would render P1 actual" and likewise for P2, and "so it does not imply the metaphysical possibility of P1 or the metaphysical possibility of P2" (loc. cit.).[1] As the last clause indicates, (ii) is the ground for (i).

[1] In her (Nida-Rümelin 2010), NR wrote: "This difference appears to be factual in this sense: 'D and Andrea is L-Andrea' and 'D and Andrea is R-Andrea' are not just two legitimate description[s] of one and the same course of events. Rather there is—according to the way we conceive of the situation—an objective *possible* feature of the world *that makes one of the two descriptions true and the other wrong*. The factual difference may be described [by] pointing out that Andrea will have a different future depending on which of the two *possible* identity facts will obtain" (2010, 196). One might be forgiven for thinking NR was assuming that P1 and P2 were possible, and that the relevant possible feature of the world, as it would make true one of the two descriptions (which here are explicitly P1 and P2), was compatible with D.

Would a factual difference in this sense help us to understand why premise 2 of the (streamlined) argument should be accepted?

I do not think that [4I] is sufficient, but I also think that there is more in the background than is expressed here, and we will come to this in a moment. The reason that [4I] is not sufficient is that a proponent of empirical criteria of personal identity over time could accept it as it stands. If one thinks that our criteria for personal identity over time involve bodily or psychological continuity (or both), then in fission cases, given that LA ≠ RA, that A is not identical with the pair of LA and RA, and that the symmetry of the case precludes choosing one over the other, the only option is to say that A does not survive. A does not survive because survival requires there be a unique closest continuer of A (which is close enough—let this be understood). Where there is no unique closest continuer, A does not survive. (This is what we say about cell fission.) How can the proponent of an empirical criterion of personal identity accept that *there is an objective feature which has to be realized in addition to what D states for P1 to be realized*? Given that it does not have to be a feature which is compatible with D, she can say that the proposition that LA is *the best* continuer of A expresses the relevant objective fact that you would have to add—though it turns out to be incompatible with D, which entails that there is no best continuer of A. Since it is incompatible with D, it is clearly something in addition to what D expresses that would have to obtain. If we were to add that the additional objective features were compatible with D, that is P1 and P2 were each possible, then this response would be closed off. But then this would be equivalent to the third proposal after all. As we will see in section 4, however, the third proposal is incoherent.[2]

---

[2] A referee for this journal suggested that one might say that the additional objective features (call them F1 and F2 respectively) were each compatible with D and that this was still compatible with P1 and P2 being metaphysically impossible. So the idea is this: that F1 obtains and D is possible, and P1 entails that F1 obtains but not vice versa, and that F2 obtains and D is possible and P2 entails that F2 obtains, but not vice versa, and that is what the factual difference comes to. But it is not required that P1 or P2 be possible. However, this divides into two cases. Either P1 or P2 is possible or neither P1 nor P2 is possible. The first case collapses into proposal three (see note 11). So we may restrict attention to the viability of the claim that F1 and F2 are both compatible with D but P1 and P2 are impossible.

This is not NR's *intended* interpretation. On the intended interpretation, the factual difference "does *not* imply that the fulfillment of description D is metaphysically compatible with the relevant feature which would render P1 actual" and likewise for P2. Therefore, as noted, on NR's interpretation of the factual difference, it could involve empirical criteria of transtemporal identity, such as being the closest (close enough) bodily continuer, since the features don't have to be compatible with D.

However, [4I] does not fully express the underlying thought. We need to look at the ground for the claim that there is an additional objective feature that would have to be added. NR says that we

---

Still, could NR appeal to this new suggestion instead? Not without rejecting the reasoning she actually engages in. I noted that (ii) is supposed to be the reason for (i).

> (i) the factual difference is compatible with the impossibility of P1 and P2,
>
> (ii) the factual difference "does *not* imply that the fulfillment of description D is metaphysically compatible with the relevant feature which would render P1 actual" and likewise for P2.

NR states (ii) and then immediately writes: "so it [the factual difference] does not imply the metaphysical possibility of P1 or the metaphysical possibility of P2." This is why she says that the factual difference is compatible with the impossibility of P1 and P2, that is, (i). More specifically, (ii) is the ground for (i) because the idea is that if only F1 could be compatibly added to D that would suffice for A = LA, and similarly for F2 and A = RA. That is why we get to say that F1/F2 are objective features that would be just what has to be added to D for P1/P2 to be realized. But if it is left open that they are not compatible with D, it is left open that P1 and P2 are not possible. If we said instead that D was compatible with F1, then, by hypothesis, the relation required for A = LA would be realized in some possible world, and so A = LA is possible. And if we said that D was compatible with F2, then the relation required for A = RA would be realized in some possible world and A = RA is possible. But if either is, surely the other is, given the symmetry of the situation, and so we would be back with interpretation 3.

Why not reject the reasoning? Couldn't one just *insist* that F1 is compatible with D (mutatis mutandis for F2) and that P1 and P2 are impossible? But now why are P1 and P2 impossible? It is not because of the symmetry involving empirical criteria of transtemporal identity, because we are rejecting empirical criteria of transtemporal identity. So if the objective non-empirical feature F1 that would be just what has to be added to D to realize P1 is compatible with D, there is a possible world in which D and the objective feature F1 which we need to add to D to realize P1 are co-realized. That is all then that is needed, given what we have said, to realize P1. So P1 is possible. Otherwise there is no content to the idea that the feature F1 is an objective feature that is what is to be added to D to realize P1. Similarly for F2 and P2.

What if one said: but why do F1 and F2 have to be sufficient for P1 and P2 (given D)? Why can't they just be *necessary*? NR aims to show that transtemporal identity of subjects of experience is not grounded by empirical criteria. Thus, in a case like the fission case except that LA/RA was the only survivor, there would be an objective feature, F1+/F2+, that was non-empirical which in the circumstances sufficed for A=LA/A=RA. Thus, F1+/F2+ is the feature that would have to be realized in addition to D for P1/P2 to be realized, where it is not merely a necessary condition but sufficient, given that D duplicates whatever empirical relations there are between A and LA/RA, in the circumstances in which F1+/F2+ are sufficient for A=LA/A=RA. And now we are back to where we started. So the trouble is this: if we want to acknowledge that P1 and P2 are both impossible, but we want to reject empirical criteria of transtemporal identity, we can't say the non-empirical condition that is all that needs to be added to underwrite P1 (mutatis mutandis for P2) is compatible with D.

> [a] … have a *clear positive understanding* of an objective feature of the world that would make it the case that one of the two successors (and not the other) is the original person. [i] If, for instance, P1 is realized, then the original person Andrea [my A] is living L-Andrea's life after the operation. [ii] Andrea has a completely different life after the operation (she has a fundamentally different future when considered from her perspective at the earlier moment) if P2 is realized. We have the conceptual capacity to positively conceive this difference … . [b] The basic idea can be put quite simply: we understand the difference between P1 and P2 by considering it from Andrea's perspective. Thinking about Andrea *in the first person mode* we are able to grasp the difference between a world in which *she* (rigid designation of Andrea) will undergo experiences related to the body with the left brain hemisphere, and a world in which *she* will undergo the experiences related to the body with the right brain hemisphere. (703-4; labeling in square brackets added)

[b] is more fundamental here than [a]. The conditionals [i] and [ii] would usually be treated as presupposing that the antecedents are possible (so that they are not just trivially true because of a necessarily false antecedent). But we are not supposed to assume that P1 and P2 are possible because D and the objective feature that would make it the case that, e.g., in the case of P1, A = LA was true may not be compatible. But given this, so far the proponent of the empirical criterion of identity can accept [i] and [ii].

Turn to [b] then. What is it to think about Andrea *in the first person mode*? To elucidate this, we should return to the original article (Nida-Rümelin 2010, sec. 4-6). This will reveal that what is crucial is the content of the positive conception of the difference between P1 and P2 (at least so far as what it excludes), and we will here identify, I believe, the fulcrum of the argument.

The key idea is brought out by first imagining oneself in A's position. You may wonder whether you will survive fissioning and whether, if you do, you will be PL (the person associated with the left hand body) or PR (the person associated with the right hand body). You wonder whether in the morning, if you wake up, you will see PL's face in the mirror or PR's face in the mirror, whether you will feel pain if PL's body is pricked, or if PR's body is pricked, and so on. You have a clear grasp, it seems, of what it would be to be one or the other of PL or PR from the first person point of view, that is, of what facts would be appropriate for the one hypothesis or the other from that point of view, where these facts can be stated in the form "I will have property *P* at moment *m*'. From this, NR concludes that

> Claim: 1: Transtemporal self-attributions are conceptually prior to self-identifications. (Nida-Rümelin 2010, 203)

The argument is given in the following passage:

> You understand the assumption 'I will be the L-person' on the basis of understanding thoughts like 'I will have property *P*.' In other words and more precisely: you understand what has to be the case for your utterance of 'I will be the L-person' to be true on the basis of your understanding of what would render your self-attribution 'I will have property *P* in the future moment *m*' true. We can formulate this claim in a more abstract way: transtemporal self-attributions (thoughts that can be expressed by sentences like 'I will have property *P*' or 'I had property *P*') are conceptually prior to self-identifications (thoughts that can be expressed by sentences of the form 'I will be *P* at moment *m*' or 'I was *P* at moment *m*'). (Nida-Rümelin 2010, 198)

It is important here also that this be a "clear *positive* understanding of what would have to be the case for [one's] thought 'I will be the L-person' to be true … due to [one's] clear *positive* understanding of what has to be the case for certain I-thoughts of the form 'I will have property P' to be true" (Nida-Rümelin 2010, 201, emphasis added).

NR adds to this a second, important claim, namely, that your "understanding of what has to be the case for your I-thought 'I will have property P' to be true in no way depends on the empirical criteria of transtemporal identity of subjects of experience that you implicitly accept" (Nida-Rümelin 2010, 199). NR argues for this by asking us to consider whether one's understanding of what it would be for one to be in pain in the future, for example, would change when one's explicit conception of the requirements for personal identity change. For example, suppose one confidently expects that one will survive when one's body and brain are destroyed provided that a new brain is created from a scan of the old with the same brain states, preserving psychological continuity. One thinks, of a moment *m* after the expected destruction of one's body, 'I will be in pain at *m*'. Then suppose one comes to reject the psychological continuity account of person identity. One comes to think now 'It is false that I will be in pain at *m*'. You change your mind about the truth of the thought 'I will be in pain at *m*' but "you have not thereby changed your understanding of the content of your own I-thought. Your conceptual grasp of what has to be the case for your I-thought to be true has not changed at all" (Nida-Rümelin 2010, 199). Only your view of what empirical criteria are necessary and sufficient for your I-thought to be true has changed. NR concludes that our conceptual grasp of what it is for one have a certain property at a future time is independent of any empirical criteria for transtemporal identity you accept.

> Claim 2: Transtemporal self-attribution is conceptually invariant with respect to changes in the thinker's accepted criteria of identity of people across time. (Nida-Rümelin 2010, 203)

This contrasts with individuals who are not conscious. Claims 1 and 2 (NR argues) entail claim 3:

> Claim 3: Transtemporal self-identification is conceptually invariant with respect to changes in a thinker's accepted criteria of identity of people across time. (Nida-Rümelin 2010, 203)

NR argues that given claims 1-3, it follows that the claims apply also for other-directed thought. If we think another is capable of first person thought, we will think that she will conceive of herself as a subject of properties in the future in a way that is conceptually fundamental relative to self-identification, and conceptually invariant with respect to views about empirical criteria for personal identity. In thinking about what it would be for another person to be one or the other of two people who result from a fission event involving her, we will think of it as a matter of what has to be so for her first person thought about her having future properties to be true. NR argues further that this extends to subjects of experience who are not capable of first person thoughts, such as infants and non-linguistic animals: we ask whether if such a subject *could* think first person thoughts, its first person thought about its future properties *would* be true, in asking about what individual it will be in the future. This then generates claims 4-6:

> Claim 4: Transtemporal attribution of properties to other experiencing subjects is conceptually invariant with respect to changes in the thinker's accepted criteria of subject identity across time.
>
> Claim 5: Transtemporal attribution of properties to others is conceptually prior to transtemporal identification with respect to others.
>
> Claim 6: The conceptual content of other-directed transtemporal identification is invariant with respect to possible changes of the accepted criteria of subject identity across time. Transtemporal criteria of subject identity do not enter the conceptual content of other-directed transtemporal identification. (Nida-Rümelin 2010, 204-205)

The two most important claims here are that in thinking in the first person mode about what our future properties are (a) we have a *positive* understanding of what is involved (b) is *independent* of our views with respect to change in the thinker's accepted criteria of subject identity across time. If this is true, then we cannot make use of the rejoinder to the

fourth proposal considered above, because while it would capture an objective fact that would be something in addition to what is expressed by D, it would not be independent of empirical criteria for transtemporal identification.

The emphasis on a positive understanding in (a) is important. It is this I wish to examine. I grant in thinking of ourselves as having properties in the future in the "first person mode" we do *not* think of ourselves under a conception that presupposes empirical criteria of transtemporal identity. But the idea that we have a positive understanding of what is involved goes beyond this, and this, as we will see, is required if the argument is to go through.

What does the emphasis on positive understanding in (a) come to and why is it important? P1 and P2 differ only in that in P1 where the predicate '= LA' appears in P2 the predicate '=RA' appears. We might say that the positive understanding of the difference involved attaches to these, but (i) this would not give any special weight to the first person mode of thought and (ii) if this is all the positive difference comes to the proponent of empirical criteria of transtemporal identity could appeal to it as well. The positive understanding that carries the weight must then, I think, attach to the mode of presentation of the self. Thus, the weight rests on the idea that we think in the first person mode of ourselves in a way that *positively* characterizes the self so as to *exclude* that the self has empirical criteria for transtemporal identity. It has to *exclude* that the self has empirical criteria of transtemporal identity, because if it leaves it open (if it were "topic neutral," for example), the argument for subject-body dualism collapses.[4]

If we add to this that

[MT] if we are identical with, composed in part, or constituted by any material object or stuff (or even immaterial stuff), then we have empirical criteria for transtemporal identity,

we can infer that we are not identical with, etc., any material object (or immaterial stuff). This would secure subject-body dualism. If we assume that necessarily subjects of experience are capable of the relevant sort of non-illusory first person thought, then we can conclude that subject-body dualism is necessarily true.

[4] A referee asked why NR would have to accept that the positive understanding excludes empirical criteria of transtemporal identity. Suppose it doesn't. Then whatever factual difference there is between P1 and P2 that rests on this positive understanding is compatible with empirical criteria of transtemporal identity, and therefore compatible with a materialist view, and therefore compatible with the rejection of subject-body dualism. So the argument for subject-body dualism won't go through if we leave it open that we have empirical criteria for transtemporal identity.

In light of this, we can see that the appeal to a *factual* difference between P1 and P2 is not essential for the argument. It rests on an assumption that will drive the argument independently. We can state the argument as follows:

1. We think (correctly) in the first person mode of ourselves in a way that positively characterizes the self so as to exclude that the self has empirical criteria for transtemporal identity.
2. If we are identical with, composed in part, or constituted by any material object or stuff (or even immaterial stuff), then we have empirical criteria for transtemporal identity. [MT]
3. We are not identical with any material object [1-2].

Call this the fundamental argument.

3. First person thought

I agree that the fundamental argument is valid, but reject the first premise. The first premise rests on claim 2. I reject claim 2, repeated here, on the reading that supports the first premise.

> Claim 2. Transtemporal self-attribution is *conceptually* invariant with respect to changes in the thinker's accepted criteria of identity of people across time. (Nida-Rümelin 2010, 203)

I accept a weaker claim, however. The weaker claim is 2*.

> Claim 2* The content of transtemporal self-attributions is invariant with respect to changes in the thinker's accepted criteria of identity of people across time.

This is distinct from Claim 2, which I take to entail, via the modifier 'conceptually', that there is a positive way in which we present the self that excludes criteria of identity of people over time. (If we say that claim 2 and 2* are the same, then the objection is that premise 1 doesn't follow.)

How can claim 2* be true though claim 2 is not? (Alternatively, how can claim 2* be true yet premise 1 not follow?) The answer is that we could think of ourselves directly, without bringing ourselves under any concept, or conception, at all, other than, perhaps, that of a thing, and think of ourselves directly as being related to a time in the future of the present by having a certain property then. This would explain why what we are thinking is invariant with respect to our criteria for transtemporal identity over time. To put it in other words, we do not pick out the future self as

> the *x* such that x exists at future time *t* and *x* bears *R* to *me* at the present time,

and attribute a property to whatever is denoted by that, but we think:

> at some time *t* in the future of the present, *I* [thought of directly at the present time] have such and such a property at *t*.[5]

To take an example, think about the proposition expressed by the English sentence 'I will have a headache tomorrow morning', as asserted by me now, taking 'I' to introduce into the proposition expressed just its referent, i.e., the speaker. The proposition is a singular proposition. The rule for the use of the pronoun involves a description:

> (R) For any *x*, any time *t*, any subject *u*, 'I' refers to *x* at *t* as uttered by *u* iff *x* = *u*.

But the rule doesn't enter into the content of the proposition, only what object it assigns as the referent of 'I'. Take '@' to be a directly referring term that picks out me.[6] Then, where '*N*' directly refers to the time of utterance (and '>' means 'is later than'), the proposition is:

> [FH] [∃*t*: *t* > *N*](@ has a headache at *t* and *t* lies within the morning of the tomorrow of *N*)

That is clearly invariant with respect to @'s views about transtemporal criteria for personal identity. If the content of the thought I have when I express myself using the sentence 'I will have a headache tomorrow morning' is given by the proposition that this sentence expresses, then while claim 2* is true, claim 2 (interpreted as sketched above) is not (because as interpreted above, claim 2 says we present ourselves in first person thought in some positive manner, as having some features, whereas if the thought about the self is direct, the self is not presented as having any features). Likewise then premise 1, repeated here, of the fundamental argument is false. What is true is premise 1′.

> 1. We think (correctly) in the first person mode of ourselves in a way that positively characterizes the self so as to exclude that the self has empirical criteria for transtemporal identity.

---

[5] I made this claim in note 3 of the 2013 paper, and related it to the final point I made in that paper. NR's reply has helped me to see that this is where the most fundamental disagreement between us lies.

[6] The reader will notice that I am not distinguishing *de re* and *de se* thoughts. I give an explanation for the substitution puzzles that motivate drawing the distinction in (Ludwig 1996). My view is that when one uses the first person pronoun, given knowledge of the rule, one uses it on the basis of thoughts that are indeed directly about the self, but when we use a proper name, even if it has the same referent, there is no guarantee that one locates the referent directly in thought. This is what gives rise to the substitution puzzles. The crucial point for the argument in the text is just that we do think of ourselves directly (in thought) when we express a thought using the first person pronoun, this is an attitude toward a singular proposition, and this expresses what we have in mind by the first person mode of thought.

1'. We think (correctly) in the first person mode of ourselves in a way that does not include that the self has empirical criteria for transtemporal identity.

1′ is compatible with our having empirical criteria for transtemporal identity. What are the implications for claims 1 and 3, repeated here?

Claim: 1: Transtemporal self-attributions are conceptually prior to self-identifications. (Nida-Rümelin 2010, 203)

Claim 3: Transtemporal self-identification is conceptually invariant with respect to changes in a thinker's accepted criteria of identify of people across time. (Nida-Rümelin 2010, 203)

In thinking about oneself as having a property in the future, as in [FH], one thinks of oneself directly at the present time, and of oneself so thought of as related by having a headache then to a time located in the future of the present time. This guarantees that the person one is thinking of as having a headache at a time subsequent to the present is oneself. But does this show that self-attributions are conceptually prior to self-identifications?

What does this mean? The natural reading is that one can self-attribute future properties without first identifying some future self as oneself. If this is what it means, then we can accept it, but it does not get us very far, if the present picture is correct. For on that picture, while it is true that when one thinks of oneself as having a property in the future, there is no question that arises about identifying oneself as the one that one is thinking about, this is just because one is thinking about oneself directly in the present as being related to a future time which one picks out by a restricted quantifier anchored by a direct reference to the time of the thought. One has not *identified* oneself as some future individual. One has only picked oneself out in the way one does when thinking a thought about a property one has at the present moment. One is not called on to think in any substantive way about what it would be for one to survive to have a property in the future. And self-attributions of future properties in this way leaves it open what would have to be true for anything at any future time to be oneself.

What about claim 3, though? This has to be given up if claim 2 is given up because it would, like claim 2, presuppose that in thinking about our future selves in the first person way we have a positive conception of the self which rules out our having empirical criteria of transtemporal identity.

These remarks carry over to claims 4-6.

So far, I have only said what follows if we accept this sketch of the content of first person thoughts about future properties of the self. One might object at this point to the account of the content of the proposition expressed by 'I will have a headache tomorrow' or to the claim that the

thought I have about properties of my future self expressed with the first person pronoun in subject position is properly specified by the proposition expressed by the sentence that I use. I respond to this in two stages. First, I give an argument to show that we do not pick ourselves out via any purely qualitative mode of presentation, so that there must be some element of direct reference in thought to the self. Second, I argue that, given this, we have no reason to think we subsume the self under any positive conception or concept in referring to ourselves in the "first person mode."

The argument for the claim that we do not pick ourselves out via any purely qualitative mode of presentation, goes as follows (Ludwig 1996).

1. We know that we are able to think about ourselves and attribute properties to ourselves.
2. If we were able to think of ourselves only by way of a purely qualitative mode of presentation (or description), then we would not know that we are able to think about ourselves and attribute properties to ourselves.
3. Therefore, we do not pick ourselves out only by way of a purely qualitative mode of presentation (or description).

The subargument for premise 2 goes as follows.

1. We do not know that the universe does not contain (timelessly speaking) qualitative duplicates of everything that exists (or at least of ourselves up to the extent of our knowledge).
2. If the universe contains qualitative duplicates of everything that exists (or at least ourselves up to the extent of our knowledge), then no purely qualitative mode of presentation (or description) uniquely denotes any individual (or ourselves to the extent of our knowledge).
3. Therefore, if we were able to think of ourselves only by way of a purely qualitative mode of presentation (or description), we would not know that we are able to think about ourselves and attribute properties to ourselves

The argument for premise 1 of the subargument is that we do not know that, for example, Nietzsche's hypothesis of eternal recurrence is not true, that is, we do not know that the universe does not repeat each temporal segment of it qualitatively identically an infinite number of times.[7]

[7] I say 'for example', because this is not the only hypothesis we could appeal to here. Do we know that there are not an infinite number of spatio-temporally isolated universes (like David Lewis "possible worlds" but without the commitment to explaining modal claims in terms of them) among which there are qualitative duplicates of ours? We could not rule these out by any empirical means, and since it is possible that a universe contain an infinite number of spatio-temporally isolated universes one of which is just like ours, there is no a priori argument to show that this hypothesis is false.

It might be said in response that physics tells us that Nietzsche's hypothesis is in fact false.[8] The second law of thermodynamics ensures that the universe will die a heat death in a state of maximum entropy. However, this overstates what we know. The matter is not entirely settled in physics whether the universe iterates through infinite cycles (a Big Bang followed by a Big Crunch, followed by a Big Bang and so on).[9] But even granting that it is settled that cycles were not physically possible—that the universe is open and will die a heat death—this doesn't matter. The fact is that we do not need to know what physicists know (if they know that) in order to know that we are able to refer to ourselves (see note 7 also). For most people, it is epistemically open that there are qualitative duplicates of them, but this is not a threat to their knowledge that they can refer to themselves. So the real force of the argument is that our knowledge that we think about ourselves is not hostage to whether there are qualitative duplicates of us, but this could be so only if we could think about ourselves directly and not only as the unique possessors of some set of (purely qualitative) properties.

This leaves it open that we think of ourselves by a mode of presentation that functions like a complex demonstrative. Let 'subject' express the special conception of the type of being that represents a positive conception of the self that excludes empirical criteria of transtemporal identity. Then it is open still that we think about ourselves in a thought of the form 'that subject is in pain'. We can think of 'that subject' as a restricted quantifier of the form 'the *x*: *x* is *that* and *x* is a subject' as suggested in (Lepore 2000), or as we could think of the function of 'subject' as a filter on how the demonstrative element refers, as Kaplan suggested (Kaplan 1989, 515). In either case, we refer directly to the self, but at the same time in doing so bring the self under the concept of a subject. The second of these components, though, does not play a crucial role in the mechanism of reference itself.

My argument against first person thought involving this extra component has three parts. First, (a) it is gratuitous and (b) we should not adopt views that are gratuitous. It is gratuitous because the main diagnostic for there being a positive conception of the self is that in self-attributing future thoughts we think of the self in a way that is invariant with respect to changes in our conception of empirical criteria for transtemporal

---

[8] Eternal recurrence does not follow form the hypothesis of infinite time. Consider two cylinders with marks that line up at time zero, one of which begins rotating once per second and the other √2 times per second. They will never line up again in the original configuration, since that would require √2 to be equal to the ratio of two integers. This is a variant of a counterexample provided by Georg Simmel in 1907 and reported in (Kaufmann 1974, 327). What is at issue is not the necessity but the epistemic possibility of eternal recurrence.

[9] See (Penrose 2010).

identity. But once we see that we refer to the self directly, we have an explanation for this that does not require that we bring the self under any positive conception. Second, we could respond to the charge that it is gratuitous if there were present in reflection on self-attributions of properties some positive conception of the self that did exclude empirical criteria of transtemporal identity. But we do not, in fact, bring ourselves under such a positive conception of the self in self-attributing properties. Or, to speak more cautiously, I do not find, that when I rap my knuckles on my desk, and have the thought that I am in pain, I am in thinking *I qua subject am in pain*, where the concept of a subject is a positive conception of what I am thinking about. It is implied by my thinking about myself at all that I am a thinking thing, but this does not enter into the content of the thought itself, apart from its "predicate". Even if it did, it would not be the right sort of conception to support the exclusion of empirical criteria for transtemporal identity. Third, if we did bring the self under a positive concept that excluded empirical criteria of transtemporal identity, then we should be able to articulate what about it precludes our having them. What is it though? Is it that we are composed of immaterial stuff rather than material stuff? This won't work because the same argument applies here, as NR notes. If we were some immaterial stuff, there would be criteria for transtemporal identity appropriate for it, but it is clear that the content of self-attributions of future properties would be invariant with respect to our conceptions of what that involved. Only thinking of the self as a thing that is utterly simple, and whose persistence through time is not governed by informative criteria at all, would seem to have the right character. This would not rule out (a priori) its being a material thing that was simple, however. And it is hardly clear that when we think about ourselves we are thinking about something that is utterly simple.[10] This seems to be an *open* question relative to our thinking of ourselves in the first person mode, in the way we express when we use the first person pronoun.

Finally, even if we did invariably bring the self under such a concept, given that it is not required for us to refer to ourselves in thought in the first person mode, it would remain an open question whether we were correct to bring ourselves under such a concept, and this would not require that we think in general that we mistakenly attribute thoughts to ourselves, for the illusion would extend only to what is from the point of view of our ordinary attributions an extraneous and unnecessary addendum to first person thoughts.

[10] This point is connected with the remark in note 10 in (Ludwig 2013) that the original argument applies only to material objects that are subject to fissioning. Absolutely simple objects are not capable of fissioning.

To summarize: in first person thought we think about ourselves directly at the present time. This is true also when we attribute to ourselves properties in the future, because we are thinking of ourselves as picked out now directly as being related to times in the future (of the time of the thought) by way of having various properties then. Thus, the content of those thoughts are invariant with respect to variations in our conceptions of criteria for transtemporal identity of people. This does not require that we have a positive concept of the self that excludes empirical criteria for transtemporal identity. The suggestion that we do bring the self under such a concept is not supported by reflection on first person thought, and in the absence of that it is gratuitous to suggest that we bring the self under such a concept. Finally, even if we did, it would not show that we did not have empirical criteria for transtemporal identity and whatever illusion this involved would be localized and not undermine the vast majority of what we think about ourselves. The fundamental argument fails.

4. Possibilities and the Third Interpretation

I turn now to the question whether the third interpretation of the factual difference between P1 and P2 supports subject-body dualism. For convenience, I repeat P1-P3 here.

P1: D and A = LA.

P2: D and A = RA.

P3: D and A ≠ LA and A ≠ RA.

The third interpretation was that P1 and P2 are both genuine possibilities. On this interpretation, the second premise of the streamlined master argument can be reformulated as in 2′.

2′. P1 and P2 could be genuine possibilities only if subject-body dualism were true.

In (Ludwig 2013), I argued (in part) that: for the possibility that P1 or the possibility that P2 to show that as a matter of fact A did not have empirical transtemporal identity conditions, it would have to be necessary that in every possible world either A = LA or A = RA; but this is incompatible with the assumption that it is possible that A ≠ RA *and* A ≠ LA, and if that last were true at the actual world, it would be compatible with our having empirical criteria of transtemporal identity in the actual world.

NR responds by introducing the assumption that we have our transtemporal criteria for identity essentially. If there is a possible world in which A = LA (when RA is also produced from A by a fission event), in that world A does not have empirical transtemporal identity conditions.

If persons have their transtemporal identity conditions essentially, it follows that in the actual world A does not have empirical transtemporal identity conditions.

Even granting the assumption, there is a problem with the third interpretation. I introduced this problem in the (2013) paper as a response to the first objection. That relied on the charge that P3 was supposed to be possible. But if it is, then, given the necessity of identity, P3 rules out P1 and P2. I replied that the rejoinder was too powerful, because it also means that P1 and P2 cannot *both* be possibilities. It would have been more straightforward, perhaps, to say that the basic problem with the third interpretation is simply that it is incoherent to maintain that P1, P2, and P3 are all possibilities. Suppose that P1 is possible. Then, given the necessity of identity, (NI),

(NI) For any x, y, if $x = y$, then for any possible world w, if x exists in w or y exists in w, then $x = y$ in w

if it is possible that A = RA, then in any world in which A or RA exist, A = RA, and since RA ≠ LA, it follows that it is not possible that A = LA. And if it is possible that A = LA, then it follows that it is not possible that A = RA. And if it is possible that A ≠ RA *and* A ≠ LA, then it follows that it is not possible that A = RA and it is not possible that A = LA. We cannot hold that each of P1, P2 and P3 are genuine possibilities. Only one of them can be. Thus, the third interpretation of the factual difference in fact involves a claim that is false, namely, that P1 and P2 are both genuine possibilities.[11]

P1, P2 and P3 strike us as prima facie possibilities, but they cannot all be genuine conceptual or metaphysical possibilities. In what sense are they possibilities? The answer, I think, and this will now connect the discussion of this interpretation with the discussion in the previous section, is that P1, P2 and P3 are epistemic, not metaphysical or conceptual possibilities. They are epistemic possibilities because when we think of ourselves in the first person mode we do so directly, without thereby revealing what our natures are.

We can imagine what it would be like to be one or the other of the two successors in a fission scenario by self-attributing future experiences appropriate for being the one or the other. We might, for example, be told that the right successor will prick the finger of her right hand while the left will prick the finger of her left hand. We can think of what it would

[11] It might be suggested that NR needs only one of the two possibilities, not both. So she might say: P1 is possible but not P2 or P2 is possible but not P1. But this is not the argument that NR is advancing, and, it seems, for good reason. For the problem with this that whatever might ground the claim that P1/P2 is possible would seem to ground the claim that the other is possible, given the symmetry of the setup. In addition, it amounts to the rejection of P3, which is the position of the materialist, and so without further argument it would be question begging.

be like to be in the position of the one, feeling the right finger pricked, rather than the left, or vice versa. In thinking of this, in the first person mode, imagining that one is feeling the prick tomorrow *there*, one thinks a thought that does not involve any incoherence. It is presented as an epistemic possibility at least relative to the content of the thought. This is not directly to think of ourselves as PR or PL. But it is a short step. For what seems compatible with D is that we are having experiences of a sort which only PR or which only PL would be having. Thus, it will seem open (relative to the content of the thought) that we are PR or PL, respectively. And when we think of others faced with fissioning, we can imaginatively project ourselves into their shoes and see that they can conceive the same thing from their point of view. It is epistemically open for them as well, relative to the content of their future directed thoughts, that they are the one or the other.

However, (a) these cannot be more than epistemic possibilities because they cannot both be metaphysical or conceptual possibilities given (NI), and (b) we have an explanation of why they are epistemic possibilities for us that does not require that they be genuine metaphysical or conceptual possibilities. For in thinking or imagining ourselves as feeling such and such a prick *there*, we have a thought of the form φ(I), where I use 'I' to represent the unmediated thinking of the subject of the thought by its subject. The nature of the object of thought is not presented in the thought. Consequently it appears open to us that it be true when all the conditions specified by the fissioning scenario are in place. But what we are picking out may be a material object even if not so presented. Consequently what we are thinking, coherently so far as the content of the thought goes, may not be a genuinely possibility, because as a matter of fact we have empirical conditions for transtemporal identity.

Our epistemic position with respect to our natures in first person thought is analogous to our position with respect to the natures of stuffs for which we introduce natural kind terms. We attach natural kind terms to natural kinds through application to examples we pick out by how they present themselves to us. But the kind properties that we aim to keep track of are not given by the features by which we pick them out. So the thoughts we entertain about them (prior to discovering what the kind property is) do not reveal their natures to us. So too in thinking of ourselves in the first person mode, how we pick out ourselves does not reveal what we are, and so it is epistemically open what sort of thing we are. The difference is that we are not picking ourselves out by any features we have, so that the self is not presented to us in the first person mode of thought in any positive way whatsoever. All of that lies in what we predicate of the self.

To summarize, NR is right that requiring that we have our transtemporal identity conditions essentially together with the assumption that, for instance, P1 is possible, entails that we are not material things or constituted from material things. But even granting the premise, this

doesn't rescue the argument on the third interpretation of *factual difference*. For the third interpretation asserts that P1 and P2 (and in fact P3) are all genuine possibilities. But this is impossible given (NI). P1, P2 and P3 are rather epistemic possibilities. That they are epistemic possibilities is explained by the fact that in first person mode future attributions of properties we pick out ourselves directly, and so in a way that does not present itself as in conflict with our being one or the other of the successors in a fission case because it is silent on what our natures are and so on whether we have empirical criteria of transtemporal identity.

5. Summary and Conclusion

I have argued that the fourth interpretation that NR offers of "factual difference" in her (2010) argument for subject-body dualism does not secure an interpretation on which the argument goes through, but that a more fundamental claim motivates this way of putting the factual difference that carries the argument by itself, namely, that first person thought involves a positive conception of the self that excludes our having empirical criteria of cross-temporal identity. In response, I argued that in first person thought, we think of ourselves directly. We do not present ourselves under a special concept that rules out our having transtemporal identity conditions. Thus, the crucial premise in the underlying argument is mistaken. I argued further that the third interpretation, that P1 and P2 are both genuine possibilities, is incompatible with the necessity of identity. They are instead epistemic possibilities that are explained by our primary reference to ourselves being direct in a way that does not reveal what sort of thing we are.

REFERENCES

Kaplan, David. 1989. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives. In *Themes From Kaplan*. New York: Oxford University Press.

Kaufmann, Walter. 1974. *Nietzsche: Philosopher, Psychologist, Antichrist*. 4th ed. Princeton: Princeton University Press.

Lepore, Ernest., & Ludwig, Kirk. (2000). The Semantics and Pragmatics of Complex Demonstratives. *Mind*, 109(434), 199-240.

Ludwig, Kirk. 1996. Singular Thought and the Cartesian Theory of Mind. *Noûs*, 30(4), 434-460.

———. 2013. The Argument for Subject-Body Dualism from Transtemporal Identity. *Philosophy and Phenomenological Research, 86*(3), 684-701.

Nida-Rümelin, Martine. 2006. *Der Blick von Innen: Zur Transtemporalen Identität Bewusstseinsfähiger Wesen*: Suhrkamp.

———. 2010. An Argument from Transtemporal Identity for Subject-Body Dualism. In *The Waning of Materialism*, edited by G. Bealer and R. Koons. Oxford: Oxford University Press.

———. 2013. The Argument for Subject Body Dualism from Transtemporal Identity Defended. *Philosophy and Phenomenological Research* 86 (3):702-714.

Penrose, Roger. 2010. *Cycles of Time: an Extraordinary New View of the Universe*. London: Bodley Head.

EuJAP | Vol. 13, No. 1, 2017
UDK
165:159.9
171:159.923

JUSTIFYING ONESELF*

MARK PIPER
James Madison University

ABSTRACT

At present, the activity of justifying oneself is mostly discussed in psychology, where it is typically viewed as a negative or at least regrettable activity involving changing one's attitudes, beliefs, and feelings in order to minimize psychological threats arising from cognitive dissonance. Yet there is conceptual space, even a need, for an analysis of justifying oneself that is more content-neutral in nature. In this paper I provide such an analysis. Along the way I also briefly canvass some of the empirical work on self-justification in psychology and gesture towards issues surrounding the normative significance of the practice of justifying oneself.

Keywords: *justification, self-justification, self-defense, commitment, internal justification, external justification, dialectical interaction*

1.Introduction

We attempt to justify ourselves constantly. Sometimes we do this casually, sometimes earnestly, sometimes desperately. We justify ourselves in response to others' challenges to our commitments and in the face of our own self-doubts. From moral exemplars to ordinary folks to flagellants, everyone feels the need to justify themselves in one way or another, perhaps more often than we would think. Justifying oneself is an ineliminable, important, even dominating feature of our lives.

But what exactly is involved in engaging in the act of justifying oneself? Note that this question is distinct from the exhaustively discussed epistemological question of what constitutes a *good* justification for a claim. The epistemological question is fundamentally normative; it concerns how justification ought to be done if one wishes to establish some kind of claim. My concern is to provide a descriptive conceptual

*Received: 10.05.2017.
Accepted: 24.08.2017.

analysis of the act of justifying oneself, which may or may not satisfy epistemic ideals. Note also that my concern is separate from epistemological inquiries into statements, propositions, or beliefs that are said to be *self-justifying* in the sense of possessing epistemic warrant without reliance on further statements, propositions, or beliefs serving as justificatory supports. By contrast, my interest is in providing an account of the *activity* of attempting to justify oneself. My analysis is thus much closer to work in moral psychology than epistemology.[1]

At present, the activity of justifying oneself is mostly discussed in psychology, where it is typically viewed as a negative or at least regrettable activity involving changing one's attitudes, beliefs, and feelings in order to minimize psychological threats arising from cognitive dissonance (Burkley and Blanton 2005).[2] Yet there is conceptual space, even a need, for a conception of justifying oneself that is more content-neutral in nature. Providing such an analysis could also be useful by clarifying its likely effects, and thereby facilitating insights regarding whether possessing a robust habit of justifying oneself tends to be productive of more benefit or harm. Although defending a normative judgment on that score would render my work more compelling in terms of practical relevance, my aspirations here are considerably more modest. I seek only to provide a theoretical analysis of the practice of justifying oneself, a practice with which we are all familiar from our daily lives. Along the way I briefly canvass some of the empirical work on self-justification in psychology, and I gesture towards issues surrounding the normative significance of justifying oneself, but my primary aim is to explore its conceptual contours. My hope is to provide an analysis that can be useful in relation to a variety of further theoretical and normative inquiries relating to this important and ubiquitous activity.

2. Empirical studies

The psychological literature is fairly united in holding the act of justifying oneself to be a negative or unfortunate activity.[3]

---

[1] The emphasis upon dialogical justificatory practice renders the present work similar in certain respects to the study of dialectical interaction, which is importantly separate from epistemology. See Rescorla (2009) for more details.

[2] Researchers in psychology distinguish between internal self-justification strategies, which involve changing one's perception of the problem behavior (by, for example, changing one's attitude towards it, trivializing it, or denying that the behavior has negative consequences) and external self-justification strategies, which involve claiming that one is not responsible for the unwanted behavior in question. See Chigona, Chigona, Ngqokelela, and Mpofu (2009).

[3] A similar, negative conception of justifying oneself can also be found amongst some philosophers. See e.g. Lang (2002) and D'Cruz (2015).

This may simply be entailed by the particular conceptions employed by psychologists, but their findings are worth mentioning insofar as they provide background for the present study and indicate the predominantly negative view amongst scholars working on the topic.

A 2008 study suggests that there is an important correlation between providing a self-serving justification and stretching the truth (Mazar, Amir, and Ariely 2008). Another study supports the claim that the prevalence of justifying oneself, understood as making excuses for negatively viewed behavior, is negatively correlated with self-esteem: the more self-esteem one possesses, the less likely one is to engage in justifying onself, and vice versa (Holland, Meertens, and van Vugt 2002). Jonathan Lowell has argued that acts of self-justification induced by cognitive dissonance can create "an amplifying feedback loop and downward spiral of immoral behavior" amongst business managers (Lowell 2012, 17). A 2014 study claims that processes of justifying oneself contribute to failures in self-regulation or autonomy (de Witt Huberts, Evers, and de Ridder 2014). According to a more recent study, increased self-justification by perpetrators of sexual aggression is a significant predictor of further sexual aggression over a one-year follow-up period (Wegner, Abbey, Pierce, Pegram, and Woerner 2015; see also Scully and Marolla 1984).[4] Another recent study argues that justifying oneself before and after intentional ethical violations tends to enable people to do wrong while feeling moral (Shalvi, Gino, Barkan, and Ayal 2015).[5]

Not all psychologists think that justifying oneself, understood here as the giving of excuses for one's behavior, is always a bad thing. C.R. Snyder and R.L. Higgins, for example, have argued that this activity can have benefits insofar as it conduces to more successful reality negotiation (Snyder and Higgins 1988). See also Kivetz and Zheng 2006.

The last citations notwithstanding, most psychologists seem to view justifying oneself as a regrettable activity, something done primarily for the sake of self-protection in the face of potentially disturbing cognitive tensions. On this view, engaging in justifying oneself goes hand in hand with some degree of closedmindedness. People who engage in justifying themselves are trying to protect the integrity of their belief systems, to avoid internal conflicts and thereby render their mental lives smooth, satisfying, and unassailable, even if that means blocking out good evidence, reasonable concerns, and the like. All too often, many psycho-

---

[4] It is instructive to note that these authors understand 'self-justification' as "making excuses for one's discrepant behavior." Given this definition, perhaps their findings aren't particularly surprising.

[5] According to this study, 'self-justification' is defined as "the process of providing reasons for questionable behaviors and making them appear less unethical." (125) For a related study, see Pittarello, Leib, Gordon-Hecker, and Shalvi (2015).

-logists worry, justifying oneself leads to the intellectual equivalent of stopping up one's ears to new ideas and perspectives until one can safely sort out why one is in the right, at least to one's satisfaction. From this point of view, justifying oneself certainly does seem to be a regrettable activity.

While it cannot be denied that justifying oneself *can* be prompted by self-serving reasons of this sort, it is far from clear that the activity as a whole deserves such general condemnation. I wish to argue, in fact, that the activity of justifying oneself in general ought not to be conceptually pigeonholed in this way. The analysis I defend below certainly makes room for unfortunate instances of justifying oneself, but it does not limit itself to them.

3. Justifying oneself

Justifying oneself is the act of defending, verbally, in writing, or in thought, the actions, values, goals, attitudes, dispositions, loyalties, intuitions, and the like (henceforth 'commitments') that one holds as part of one's identity in response to an external or internal challenge, in the hopes, at best, of bringing the challenger to accept the commitments they have challenged, or, at the least, of resulting in a conviction in the one justifying that the demands of justification have been met and hence that no further justification is necessary. Eight aspects of this analysis require further comment. In what follows, 'justifying oneself' should be understood as shorthand for 'the act of justifying oneself'.

(1) Notice first that justifying oneself encompasses attempts to justify a person's commitments in the broadest sense, not only the person's personality. While justifying oneself may seem to relate only to general attacks on one's character and the like, this is too narrow. Our commitments are parts of who we are. When we defend parts of ourselves that are sincerely bound up in our identities – our ideas, suggestions, values, attitudes, actions, nonactions, interpretations, traits, thoughts, policies, goals, and so on – we are justifying ourselves.

(2) Justifying oneself is primarily an act of defense. This is important insofar as it establishes that justifying oneself conceptually requires a prior moment of challenge. Yet the challenge shouldn't necessarily be conceived as an attack. Although some challenges are robust or offensive enough to warrant that description, many challenges are little more than passing doubts or friendly questions that potentially call for a justificatory response.

(3) Contrary to the dominant view amongst psychologists, the act of defense that characterizes justifying oneself needn't be motivated by the desire to protect the integrity of one's belief system. Justifying oneself

needn't always be blindly or closedmindedly self-protective. The resistance constitutive of justifying oneself may just as easily be motivated by the correct conviction that one has excellent reasons to have the commitments one has. In short, justifying oneself can be and often is prompted by a correct commitment to reason rather than by a desire to protect oneself at all costs in order to avoid cognitive dissonance.

(4) In a similar vein, justifying oneself does not imply closedmindedness. It may be thought that justifying oneself implies that the one engaging in it has already made up her mind on the issue. After all, why would one justify one's commitments unless one thought that those commitments are correct or appropriate? But while some cases of justifying oneself may involve closedmindedness, it is not difficult to imagine cases in which a person engages in justifying oneself but remains open to the possibility that he is mistaken. People often claim to have beliefs but be humble about them. While some people merely say this but don't mean it, many people say it and mean it. Many cases of justifying oneself can thus be viewed as *provisional* in the sense that the one engaging in the justification thinks that it is warranted at the time or is at least committed to pushing the warrant at the moment of justification, but remains open to the possibility that counterevidence might lead one to a change of mind, possibly even soon after providing the justification. Reflection on everyday experience suggests that this is not at all uncommon.

(5) Justifying oneself only takes place in response to challenges to our sincerely-held commitments. By this I mean commitments that we consider important in some way for our identity or person, commitments that are typically bound up with emotional attachments and the threat of personal loss of some kind. Challenges that do not touch a 'self'-nerve in this way might call for response of some kind, but it is misleading to characterize them as acts of justifying oneself. If I suggest that we should paint the wall blue, but I don't really care, and you say that we should paint it green, I am not thereby pulled into an act of justifying myself; I can shrug and agree without any feeling of personal loss. If, however, I am sincerely attached to the idea of painting the wall blue, for whatever reason, and you challenge this idea, then justifying myself is on the table again. Obviously there will be grey areas when it comes to distinguishing between challenges that do not call for justifying oneself and those that do, but I will not attempt to defend criteria by which such muddy cases can be adjudicated.

(6) A challenge needn't be given by an external source. It could be provided by oneself and for oneself in the course of reflecting upon one's commitments. Moreover, in cases of internal critique, the challenge needn't be consciously formulated. On the plausible assumption, which I will not defend here, that subconscious challenges exist and influence us, it is possible that a person may feel compelled to justify her commitments

to herself out of an inchoate apprehension that they stand in need of further support or validation.

(7) As mentioned, an act of justifying oneself has a purpose: at best, to bring the challenger to accept the challenged commitments, or, at the least, to bring the one attempting the justification to the belief that the challenged commitments have been defended well enough such that no further justification is necessary. The latter goal is in some ways the more primary; for even if a person engaged in justifying herself fails to convince her objector, she will have successfully performed an act of justifying herself if *she* considers her defense to have been adequate. Justifying oneself thus inherently involves providing a defensive response to a challenge of some kind in the hope of establishing, for oneself or others, that one's commitments are acceptable in some sense. This is why there are close relations between justifying oneself and the feeling in the one challenged of vindication – the feeling of, 'on the contrary, this is acceptable!'

The word 'acceptable' may strike some as unhelpfully imprecise, but I have chosen it with care for the reason that one may engage in justifying oneself with the goal of eliciting a variety of different reactions of acceptance from one's challenger. These reactions can be viewed as lying on a spectrum characterized by differing levels of acceptability. Starting from the highest levels and working downwards, the scale would look something like this:

Faultless

Excellent

Good

Reasonable

Satisfactory

Tolerable

No worse than anything else

Not entirely daft

The hoped-for level of acceptance in any given situation will depend upon the particulars of the situation. If I am defending a paper at a conference and one of my arguments is challenged, I may attempt to justify my commitments with such strength and clarity that my objector comes to realize that his challenge was entirely inadequate for some reason, such as failing to note a caveat I had made, failing to see a relation of implication, failing to see that his argument is based upon false premises, and the like. In this case, assuming my defense is sound, my objector may withdraw his objection and judge my justification to be faultless or excellent. A student justifying her decision to go to Panama City for spring break in response to his parents' worries about his safety

would likely not expect such success; he might hope to justify his decision in such a way that his parents find his justification good or reasonable or satisfactory or perhaps simply tolerable. A president called by a leader of the political opposition to defend her foreign policy decisions on the grounds that they will lead to catastrophe may, in her act of justification, only hope to establish in her objector's mind, or in her own, if she still entertains doubts, that the policy is no worse than any other feasible policy, or that the policy, even if imperfect in some ways, is not entirely daft.[6]

(8) Note that the kind of acceptance discussed above is not to be understood as necessarily satisfying epistemic desiderata. The goal of epistemic justification is truth, and epistemic demands may be satisfied without any psychological acceptance taking place. Successful justification of oneself as here conceived amounts to a form of psychological *closure*: even if only for the moment, the challenger is silenced or, at the very least, the one justifying oneself is satisfied that the demands of justification have been met. Unfortunately, from an epistemic point of view at least, this can happen with or without rational reasons in support of that experience of closure. Successful attempts to justify oneself, in the broad sense under discussion here, can thus include or not include actual epistemic success. For this reason, an act of justifying oneself should not only be conceived as giving a well-structured argument with supporting evidence in response to a challenge. There are many ways to attempt to defend a commitment, some of which have better rational credentials than others, but all of which can function to put lingering doubts to rest, if only for a time, in order to achieve the desired acceptance. Well-structured arguments can certainly play this role, but explanations, clarifications, consideration of new or extenuating factors, and even rhetorical tricks can be employed in the service of this end. Thus we can say in one case that although someone justified a belief to her satisfaction in the sense discussed above, her justification was not in fact rational; but we could just as well say in another case that someone not only justified a belief to his or another's satisfaction, but did so in a way that satisfies rational desiderata as well.[7]

4. Relation to associated theoretical issues

(1) Justifying oneself as understood here is entirely normatively neutral regarding the content that it seeks to justify. There are no normative constraints upon the objects of justification. This is why some cases of

---

[6] My thanks to an anonymous reviewer for the *European Journal of Analytic Philosophy* for suggesting this further level of acceptance.

[7] I would like to thank an anonymous reviewer for the *European Journal of Analytic Philosophy* for pressing me to elaborate on these points.

justifying oneself will be quite faultless in all normative, epistemic, or logical respects. Other cases will involve error in one or more of these respects.

(2) Cases of positive evaluation of oneself without relation to a prior moment of challenge do not count as justifying oneself. If, absent any form of challenge to one of my commitments, I reflect upon it and find it good in some way, I am not engaged in an act of justifying myself. When I reflectively review and applaud my commitments without prior prompting by any kind of challenge, I could be said to be instead engaged in an act of endorsement. Unlike justifying oneself, endorsement without prior challenge is entirely coherent. Obviously there are muddy waters here. Are we sure that there are no lurking subconscious challenge-prompts? Are we sure that we are not reflectively reviewing our commitments because of a now-forgotten challenge sourced in a past conversation or reading? Often it will be hard or even impossible to tell. But on the assumption that no such challenges are present, yet an act of endorsement takes place, the endorsement should not be confused with an act of justifying oneself.

(3) Justifying oneself has conceptual similarities to giving excuses, but there is at least one important difference. An excuse is given by or on behalf of a person in response to a claim of wrongful action or inaction in order to abrogate responsibility for that action or inaction. Excuses are thus inherently linked to evading or repudiating responsibility for something. By contrast, it is entirely possible to justify oneself without seeking to annual responsibility for one's commitments or actions. We find a good example of this in Plato's *Crito*.

(4) Justifying oneself has conceptual links to adaptive preferences, but they are not identical. Adaptive preferences are preferences for future things, where those preferences have been formed in response to oppressive or unfavorable past conditions (Elster 1989; Christman 2014; Stoljar 2014). Justifying oneself involves justifying existing commitments, which need not have been formed by oppressive or unfavorable past conditions, but which have been called into question by some form of prior challenge. Many cases of justifying oneself are entirely appropriate, benign, and reasonable. Many cases of justifying oneself have superb rational credentials and take place without the causal history distinctive of the formation of adaptive preferences. Of course, some cases of justifying oneself are closely related to adaptive preference formation. I may be challenged on an adaptively-formed, sincerely held commitment and thereby called to engage in an act of justifying myself. Yet I can just as well have adaptive preferences that are never challenged, or I can have adaptive preferences that are challenged but are not sincerely held by me, or I can attempt to justify commitments that are not adaptive preferences.

(5) Justifying oneself is also conceptually close to wishful thinking. Wishful thinking is characterized by forming beliefs and making decisions based upon what we would like to be true or false rather than upon what evidence or careful argumentation supports. Wishful thinking, which constitutes an informal logical fallacy when it is used to defend beliefs, may be a way in which to engage in justifying oneself, admittedly a poor one from a rational perspective, and perhaps this is not altogether uncommon. But it is quite possible to engage in justifying oneself without using wishful thinking methods, and it is also possible to engage in wishful thinking when forming beliefs or making decisions in a way that doesn't involve justifying oneself. This would happen when there was no prior challenge to a person's sincerely-held commitments, for example. Other differences are notable as well. Some cases of justifying oneself can be epistemically respectable, whereas justification by wishful thinking is always fallacious. And while both justifying oneself and wishful thinking can be prompted by a perceived challenge to one's commitments, the threat in the wishful thinking case doesn't have to be a challenge. It could simply be prompted by the unfortunate circumstances in which one finds oneself (e.g. the grapes are out of reach).

(6) We can distinguish among self-justifying acts, habits, and people. Self-justifying acts are one-off attempts at justifying oneself. A self-justifying person is a person who has a regular and wide-ranging habit of engaging in self-justifying acts. Self-justifying people needn't be understood as always and everywhere engaging in this activity, only regularly and widely. Someone who lacks this habit may certainly engage in acts of justifying herself on this or that occasion, provided that such acts are not grounded in an entrenched, pervasive disposition.[8] Attempting to provide criteria to underwrite a sharp distinction between the two types of person would take us too far afield, and would likely be fruitless anyway.

(7) Robustly self-justifying people have a habit of attempting to justify most of their commitments most of the time in response to a wide range of perceived challenges, internal and external, imagined and real. Such a habit can have both beneficial and harmful effects. Although I will not provide arguments for the following claims, I would speculate that the likely *positive* effects of a habit of justifying oneself would include, in different circumstances and for different characters, (a) a sense of self-esteem/self-love/self-worth, (b) a sense of belonging, (c) self-knowledge/self-awareness, (d) autonomy, (e) self-assurance/self-confidence, (f) rightful self-promotion, (g) a sense of contentment, (h) a habit of self-scrutiny, (i) taking oneself seriously, and (j) taking others' views seriously. The likely *negative* effects would include, in different circum-

[8] For an analogous analysis that relates to an autonomous act (local autonomy) vs. an autonomous person (global autonomy), see Christman (2015).

-stances and for different characters, (a) closedmindedness/self-blindness, (b) excessive defensiveness, (c) self-deception, (d) selfishness, (e) immorality, (f) excessive self-involvement/self-indulgence/self-absorption, (g) excessive self-flattery/self-importance/self-promotion, (h) unjustified righteousness, (i) deepening of various cognitive biases, (j) denial of responsibility, and (k) the precluding of possibilities for growth.

5. Final comments

In this paper I have provided a conceptual analysis of a broad and normatively content-neutral conception of justifying oneself. The conception I have defended balances out the predominantly negative conception of justifying oneself employed by most psychologists and reopens speculation as to the normative significance of possessing habits of justifying oneself.

A variety of theoretical and normative inquiries relating to this pervasive activity might be pursued. Do challenged commitments in fact need to be sincerely held in order for a response to be considered a case of justifying oneself? Can we draw a line between healthy and unhealthy engagement in justifying oneself in response to internal critique, and if so, how? Can we speak of ideal forms, or habits, of justifying oneself that cannot be fully unpacked in epistemological terms? If so, what do these look like, and how might education for that ideal be structured? Are habits of justifying oneself in fact reliably correlated with the positive and negative effects discussed above? Is it possible to make *cateris paribus* judgments about the general value of the habit of justifying oneself, or are all such judgments ultimately a matter for casuistry? I have not been concerned to provide answers to these or related questions. My goal has rather been to trace the basic conceptual contours of this important and little discussed phenomenon in the hope that it might aid in such inquiries.

REFERENCES

Burkley, M. and H. Blanton. 2005. When Am I My Group? Self-Enhancement Versus Self-Justification Accounts of Perceived Prototypicality. *Social Justice Research* 18: 445-463.

Chigona, W., A. Chigona, B. Ngqokelela, and S. Mpofu. 2009. MXIT: Uses, Perceptions and Self-Justifications. *Journal of Information, Information Technology, and Organizations* 4: 1-16.

Christman, J. 2014. Coping or Oppression: Autonomy and Adaptation to Circumstance. In *Autonomy, Oppression, and Gender*, eds. A. Veltman and M. Piper, 201-226. Oxford: Oxford University Press.

Christman, J. 2015. Autonomy in Moral and Political Philosophy. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed May 30, 2017. http://plato.stanford.edu/archives/spr2015/entries/autonomy-moral/.

D'Cruz, J. 2015. Rationalization as Performative Pretense. *Philosophical Psychology,* 28: 980–1000.

de Witt Huberts, J.C., C. Evers, and D. T. D. de Ridder. 2014. 'Because I Am Worth It': A Theoretical Framework and Empirical Review of a Justification-Based Account of Self-Regulation Failure. *Personality and Social Psychology Review* 18: 119-138.

Elster, J. 1989. Sour Grapes – Utilitarianism and the Genesis of Wants. In *The Inner Citadel*, ed. J. Christman, 170-188. New York: Oxford University Press.

Holland, R.W., R. M. Meertens, and M. van Vugt. 2002. Dissonance on the Road: Self-Esteem as a Moderator of Internal and External Self-Justification Strategies. *Personality and Social Psychology Bulletin* 28: 1713-1724.

Kivetz, R. and Y. Zheng. 2006. Determinants of Justification and Self-Control. *Journal of Experimental Psychology: General* 135: 572-587.

Lang, B. 2002. Moral Clichés (or, How Not to Teach Ethics). *Teaching Philosophy* 25, no. 3 (Sept 2002): 247-250.

Lowell, J. 2012. Managers and Moral Dissonance: Self-Justification as a Big Threat to Ethical Management? *Journal of Business Ethics* 105: 17-25.

Mazar, N., O. Amir, and D. Ariely. 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45: 633-644.

Pittarello, A., M. Leib, T. Gordon-Hecker, and S. Shalvi. 2015. Justifications Shape Ethical Blind Spots. *Psychological Science* 26: 794-804.

Rescorla, M. 2009. Epistemic and Dialectical Regress. *Australasian Journal of Philosophy* 87: 43-60.

Scully, D., and J. Marolla. 1984. Convicted Rapists' Vocabulary of Motive: Excuses and Justifications. *Social Problems* 31: 530-544.

Shalvi, S., F. Gino, R. Barkan, and S. Ayal. 2015. Self-Serving Justifications: Doing Wrong and Feeling Moral. *Current Directions in Psychological Science* 24: 125-130.

Snyder, C.R. and R. L. Higgins. 1988. Excuses: Their Effective Role in the Negotiation of Reality. *Psychological Bulletin* 104: 23-35.

Stoljar, N. 2014. Autonomy and Adaptive Preference Formation. In *Autonomy, Oppression, and Gender*, eds. A. Veltman and M. Piper, 227-252. Oxford: Oxford University Press.

Wegner, R., A. Abbey, J. Pierce, S. E. Pegram, and J. Woerner. 2015. Sexual Assault Perpetrators' Justifications for Their Actions: Relations to Rape Supportive Attitudes, Incident Characteristics, and Future Perpetration. *Violence Against Women* 21: 1018-1037.

EuJAP | Vol. 13, No. 1, 2017
UDK
165.172
161.25

A DEFENCE OF A RATIONALIST CONCEPTION OF PRACTICAL REASON*

GAL YEHEZKEL
The Sapir Academic College & The Open University of Israel

ABSTRACT

In this paper I attempt to refute the instrumental conception of practical reason, and thus defend a rationalist conception of practical reason. I argue that, far from merely playing an instrumental role, reason can be used by an agent to evaluate, that is, to approve or reject, final ends, which might be suggested by desires, and further to determine final ends independently of any desires, whether actual or potential, that the agent might have. My argument relies on an analysis of the concept of intention, and, more specifically, on the distinction between want and intention. I argue that the notion of an intentional action entails that reason can be used to evaluate and determine final ends.

Keywords: *end, instrumental reason, intention, practical reason, reason*

1.Introduction

Can agents rely on reason alone to determine their final ends? Can agents even use reason to evaluate the ends they set for themselves? According to an instrumental conception of practical reason, reason can *only* serve an instrumental role. That is, reason can only be used by an agent to determine means (or instrumental ends), but not to determine or evaluate their final ends. Thus, final ends, which are the things we pursue for their own sake, and terminate the chain of justification, cannot themselves be rationally justified. Hence this view is sometimes called "subjectivism" in order to emphasize the idea that practical reasons are agent-relative, as they are derived from the agent's own subjective, contingent, conative states.

The instrumental conception of practical reason is most famously attributed to David Hume. There is some controversy as to the specifics

*Received: 10.05.2017.
Accepted: 24.08.2017.

of Hume's position (see, for example, Hampton 1995; Weller 2013). However, my aim in this essay is not exegetic. I am concerned with the view that is expressed by the well-known quote from *A Treatise of Human Nature*: "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (Hume 1978, 415). My aim is to show that this view is false.

Here, by contrast, I intend to defend a rationalist conception of practical reason. I argue that, far from playing a merely instrumental role, reason can be used by an agent both to evaluate, that is, to approve or reject, final ends, which might be suggested by the agent's desires, and to further determine final ends independently of any desires, whether actual or potential, that the agent might have.

The strategy I employ involves exposing the shortcomings of its rival conception, that is, the instrumental conception of practical reason, mentioned above. Two (related) explanatory notes are in order. To begin with, I do not assume that instrumentalism and rationalism are the only two ways of conceiving practical reason. However, my criticism of the instrumental conception of practical reason inevitably leads to a rationalist conception of practical reason. Furthermore, it might be argued that my criticism is directed towards pure instrumentalism, that is, the view that reason can only be used in order to determine means to desired ends. There are however advocates of the Hume-inspired subjectivist school of thought who argue against pure instrumentalism, and maintain that reason serves a more complicated role than pure instrumentalism recognizes (see, for example, Schmidtz 2001; Reitsma 2013). Hence, it might be argued that arguments against pure instrumentalism do not refute the subjectivist school of thought. In response, I should note that I do not attempt to infer the rationalist view of practical reason from the falsehood of pure instrumentalism. Rather, pure instrumentalism, as a minimalist view of the role of reason, is the starting point of my analysis. In the course of my analysis, the true role of reason in practical reason is revealed, and a rationalist view is established.

Criticism of the instrumental conception of practical reason is not new (see, for example, Hampton 1998; Korsgaard 1998; Lebar 2004). However, there are many who believe that this criticism ultimately fails to disprove this view (see, for example, Hubin 2001; Andreou 2005; Spielthenner 2012; Markos 2014). Although I see merits in previous attempts to refute the instrumental conception of practical reason, in this paper I wish to advance a novel argument to this conclusion. I believe that my argument has the merit of not only refuting the instrumental conception of practical reason, but also of shedding light on the intricate relations between key concepts in practical reasoning, that is, desires, wants, intentions, actions, and ends.

Hence, in this essay I argue that we can use reason not only to evaluate

ends that might be suggested by desires, but also to determine the final ends an agent should pursue independently of any desires that the agent might have, actual or potential. My argument does not rely on any contingent premises; for example, that individual agents (usually) have more than one end (Hampton 1998, 167). That is, I do not argue that it is only a matter of contingent fact that it is possible for reason to determine final ends. I argue that the notion of an intentional action entails that reason can be used both to evaluate and determine final ends. More specifically, I argue that the concept of intention entails a distinction between what is wanted and what is intended, and since reason is responsible for this distinction, it is the use of reason that enables an agent to determine his or her intentions and hence his or her ends.

I should stress that my aim in this paper is not to develop or defend a *specific* rationalist theory of practical reason (for example, that which forms the heart of Kant's moral theory). Hence, I do not attempt to identify any specific final ends that a rational agent should pursue.

Furthermore, rather than defending a rationalist conception of *motivation*, according to which reason alone can motivate actions, my interest is in defending a rationalist conception of practical reason. I argue that reason *alone* can be used to *determine* the final ends that an agent *should* pursue. Prima facie, it seems possible to integrate a Humean Theory of Motivation, according to which "reason alone *can never be a motive* to any action of the will" (Hume 1978, 413), with a rationalist conception of practical reason, according to which reason alone *can be used to determine* which final ends the agent should pursue (see, for example, Smith 1988). Although attempts to integrate these views have been criticized (see, for example, van Roojen 2002), in this essay I do not presume it to be impossible.

In section two I analyze the concept of intention. I focus on the relations between the concepts of intention, desire, want, action, and explore the role of reason in explaining the distinctions between these concepts. I show that the notion of intention, and hence the notion of an intentional action, relies on the distinction between a want and an intention. My main argument, which is developed in the following sections, is that this distinction, between what an agent wants and what he or she intends, relies on the use of reason to evaluate ends. In section three I discuss cases in which an agent has more than one end. Such cases have been used before in criticism of the instrumental conception of practical reason. However, my conclusion is that although instrumentalists must admit that the use of reason is necessary to explain the distinction between want and intention in these cases, they ultimately fail to disprove this view. Hence, in section four I turn my attention to a hypothetical case in which an agent has only one desire and only one end. I show that the distinction between a want and a desire is applicable to this case, and that this can only be explained by the use of reason. My analysis shows that

the concept of intention entails that reason can be used to evaluate, that is, to approve or reject ends, which might be suggested by the agent's desires. In section five I argue that reason can not only evaluate ends that might be suggested by desires, but can also be used to determine ends that are not suggested by an agent's actual or even potential desires. Hence, reason has the authority to determine final ends altogether independently of desire. In section six I summarize the conclusions of my analysis.

2. Terminology and the distinctions between desire, want, and intention

I begin my discussion by analyzing the connections between the concepts of *end*, *action, intention, want,* and *desire*. This not only clarifies the connections between these concepts, but most importantly illuminates the differences between them.

Before turning to analyzing the relationships between these concepts, I concede that it is impossible to make a claim about them that has not yet been rejected in the literature. Nevertheless, there are claims that are generally agreed upon, and claims that are more controversial. In what follows I attempt to outline important distinctions between different aspects of human actions. First and foremost I argue that these conceptual distinctions exist and try to explain their importance. The terminology I use to describe these distinctions is of secondary importance. I try to use terms in accordance with their pre-theoretical use and defend the more contested choices of terminology I make. However, I do not expect my use of these terms to fully conform to their pre-theoretical use, or satisfy every theoretician. The most important aspect of my analysis is the distinctions that are represented by these terms, rather than the choice of terminology. Furthermore, my main focus in this analysis is the easier task of exposing conceptual *distinctions*, rather than conceptual ties. For in many cases a simple example can demonstrate the distinction between two concepts, while a claim for a connection of entailment, as a general connection, cannot be merely demonstrated and should always be argued for.

Let’s begin by considering the following propositions:

(a) Naomi desires to eat ice cream.

(b) Naomi wants to eat ice cream.

(c) Naomi intends to eat ice cream.

(d) Naomi eats ice cream.

(e) Naomi's end is to eat ice cream.

Beginning with the concept of action, as expressed, for example, in (d),

there are several distinctions that one can make. The first is between intentional and unintentional actions or mere behavior (Anscombe 1957, 84). Although one might doubt the existence of unintentional actions, for the present discussion it is unnecessary to enter into this debate, because our interest is in intentional actions, and what separates them from mere behavior. (Hence, from now on I will simply use the term "action" as an abbreviation for intentional action.) Furthermore, this essay assumes that an intentional action entails intention on the part of the agent (Davidson 1963; Chisholm 1964). This view has been criticized (for example, Anscombe 1957; Bratman 1984). However, I believe that the objections to this view have been sufficiently answered (Wasserman 2011). I should however stress that my argument does not assume the more controversial claim that for each thing a person does intentionally she intends it under the description used to attribute the intention. I merely assume that an intentional action requires the ability to intend to act, that is, if the concept of intention cannot be meaningfully attributed to an agent, it is meaningless to ascribe an intentional action to this agent.

Intention is always directed towards an action, which is supposed to bring about a wanted outcome, that is, an end. This is true both in cases in which that action is a final end and in cases in which it is only a means for bringing about another (perhaps final) end. Hence, an agent cannot intend to "world peace," but can intend to *bring about* world peace.

Some writers have made several distinctions between different categories of ends. By an "end" I simply mean a final end, that is, something we pursue for its own sake – our goal, rather than an instrumental end, a constitutive end, or a maieutic end (see Schmidtz 2001, 238-239). An end for an agent is something that the agent intends to bring about. Therefore, proposition (e) entails proposition (c). If the agent merely wants something, for example, "world peace," but does not intend to do anything to bring it about, it is not an end for this agent. (The converse does not hold true, however: proposition (c) does not entail proposition (e); not everything an agent intends to bring about is an end for him or her, and may only be a means to an end.)

Notice that even if one assumes that proposition (d), which describes an intentional action, implies the existence of an intention to eat ice cream (proposition (c)), the distinction between intentional action and intention is nevertheless maintained. This is because the converse obviously does not hold true. That is, the existence of an intention, as a mental attribute, does not entail the existence of an action, which is possibly a physical event that is supposed to be caused by that mental state. Thus, the possibility of an intention without an action ensures the conceptual distinction between intention and action. (Obviously I use the term intention in the sense of past intention and not in the sense of “intention in action” (Searle 1983, 83-98).)

Intentional action hence entails the possibility of intention, and depends upon it. An analysis of the conceptual structure supporting intention is therefore necessary in order to understand the concept of an action. In what follows I focus on the distinction between intention and want, and examine the necessary conditions for maintaining this distinction.

However, before turning to the distinction between intention and want, I wish to clarify my use of the term "want," and the distinction I find important between "desire" and "want."

In proposition (a), eating ice cream is identified as something the agent desires. As previously mentioned, according to an instrumental conception of practical reason, reason can *only* be used by an agent in an instrumental role, while final ends are determined by desires. Obviously, in common use the term "desire" is too restrictive and narrow to explain all the final ends that agents have (Searle 2001, 167–170). Having a desire, in the sense of the term "desire" that follows this common usage, is just one way in which our final ends are determined by our desires or subjective contingent conative states. From now on I will use the term "desire" to indicate any subjective contingent conative state that can determine, at least according to the instrumental conception of practical reason, our final ends. Furthermore, these final ends are not subject to rational constraints, and, hence, are neither rational nor irrational. (My use of this term can perhaps be likened to what Hubin (1991) terms a "basic, unmotivated desire," in contrast to a "motivated desire".)

Since Naomi desires to eat ice cream, it is something that Naomi *wants* to do. Hence, proposition (a) entails proposition (b). A desire entails the existence of a 'want.' This, however, might be obscured by the fact that sometimes a desire is defeated, for example, by a stronger desire, and a matching want. In this case it might be the desire to keep one's figure. However, even if an agent decides not to satisfy a desire to eat ice cream, the want to eat ice cream still exists.

Obviously, Naomi might answer the question "Do you want to eat ice cream?" with the reply "No," but this is only the case if the question is interpreted as an offer of ice cream (for example, if a host is offering his or her guests ice cream for dessert). She might also answer the question, "I want to – actually I have been craving ice cream all day – but I also want to stick to my diet, so no thanks." This possibility shows that when we deny the existence of a want, where a desire clearly exists, we do so only in order to avoid a misunderstanding as to whether we are intending to act according to our want (see also Audi 1973, 7). The existence of a desire (a) therefore always entails the existence of a want (b), although this want can be defeated by a conflicting, stronger want.

The alleged connection between desire and want, that is, that desire (a) implies a want (b), may be contested, but I wish to stress that although it is clearly important to understand the correct relations between the basic

concepts I am analyzing, the main argument of this paper does not rely on this claim. On the contrary, as I explain in what follows, I believe that if a desire (a) does not entail a want (b), this fact alone undermines the instrumental conception of practical reason, and in fact can be integrated with the main argument of this paper.

In order to understand why, let us begin by noting that even if a desire (a) does not entail a want (b), there are clear cases in which a desire does determine a want. This is evident from the meaningfulness of the claim "The baby wants to eat." The fact that a baby can be meaningfully said to want to eat shows that a desire can directly determine a want, because in this case it is clear that the baby's want is not determined by reason. Hence, if one insists that a desire (a) does not entail a want (b), one can only turn to reason in order to explain the possibility of a desire that is not accompanied by a matching want.

It might be objected that this possibility can be explained by conflicting desires. In this case, it might be argued, the stronger desire determines what the agent wants, while the weaker desire is dismissed. However, in order to relinquish his or her original want the agent must *acknowledge this conflict*, and this can only be explained by the use of reason. Hence, even according to this suggestion, only reason can determine our wants, and hence reason acts as a critic of desires. That is, reason can prevent desires from determining our wants. This conclusion implies that, contrary to the instrumental conception of practical reason, desire alone determines neither what we want, nor (as I argue in what follows) our intentions and our actions. Due to the fact that intentional actions are directed towards ends, if reason is used in order to determine our intentions, it is also used in order to determine our ends. Thus, the idea that a desire (a) does not entail a want (b) can be used to refute the instrumental conception of practical reason.

Again, I wish to stress that *I do not pursue this line of argument* simply because I believe that a desire (a) does imply a matching want (b). I present this consideration in order to make clear that denying this implication does not disprove the conclusion of this paper.

Turning back to the relation between desiring and wanting, a basic difference between a desire and a want should be noted. A desire is non-rational, in the sense that it is meaningless to ask, "Why do you desire to eat ice cream?" It is always meaningful on the other hand to ask, "Why do you want to eat ice cream?" Although an answer to the latter question can be given by referring to a desire to eat ice cream, the question and the answer are not trivial, because it is possible that eating ice cream is merely a means to an end, and hence determined with the help of reason. This demonstrates the connection between reason and wanting.

The connection between reason and wanting does not imply that every want is determined by reason, as becomes evident when we observe the

meaningfulness of the claim "The baby wants to eat." It does show, however, that we find a use in language for the term "want," as distinct from that of "desire," because of the possibility for a certain want to be determined by the use of reason, rather than by a desire. That is, the distinction between desiring and wanting is maintained by the fact that although a desire (a) entails a want (b), the converse does not hold true. An agent wants what he or she desires, but also wants things that he or she does not necessarily desire. This is possible in the case in which an agent wants something only as a means to an end. For example, even though Naomi hates the taste of ice cream and therefore has no *desire* to it eat, she may *want* to eat it, based on her dentist's recommendation to eat ice cream following a dental extraction. Reason, in this case, is used to determine means, which is not something which is governed by desire. The distinction between desire and want is therefore supported by the fact that the concept of a want is broader than the concept of a desire.

The logical distinction between desire (a) and want (b) therefore reflects the use of reason in determining means to final ends. In this use, reason must enable the agent to determine a want to bring about a means to an end, although there is no desire to bring about this means. So far the conclusions of our analysis are in accordance with the instrumental conception of practical reason. However, according to the instrumental conception of practical reason this is the only use of reason in practical reasoning. In order to see whether this view is correct let us now turn our attention to the distinction between want (b) and intention (c).

An agent's intention to perform an action, as described in proposition (c), entails that the agent 'wants' to perform this action, as described in proposition (b) (see also Schueler 1995, 35). Although some reject this claim, it is evident by the fact that it is always legitimate to ask "Why do you *want* to do it?" when an agent declares his or her intention to do something (Thompson 2008, 104).

Obviously the agent may not find any intrinsic value in a specific action, and only identify it as a means to a final end. However, if the agent intends to perform this action (even under threat), he or she nevertheless wants it to take place under these circumstances. There is obviously a sense in which an agent could consistently say that she intends to go to the dentist although she does not want to. However, what the agent would mean by this claim is that she does not want to go to the dentist in and of itself, but only wants to do it all things considered, that is, only as a means to stopping the pain she feels. It is of course possible to go to the dentist without wanting to do so. This is the case, for example, if I want to go an optometrist but have the wrong address, and end up at the dentist. However, in this case it would be false to say that I went *to the dentist* intentionally.

Although an intention (c) entails a want (b), the converse does not hold

true. An agent may (desire and hence) *want* to eat ice cream, but not *intend* to eat ice cream, due to dietary restrictions. Hence, a want does not entail an intention to act according to the want. Reason can intervene, and the agent may therefore decide not to act on a certain want. The fact that proposition (b) does not entail proposition (c) therefore supports the distinction between the concept of want and the concept of intention.

Notice that the divergence between want and intention can only apply to what is wanted as a final end, and not to what is wanted as a means. It is possible, for example, for reason to intervene in a situation in which a desire determines a final end, while reason can exclude any intention to bring about the desired end (I discuss this possibility in detail below). A means, on the other hand, is determined by the use of reason. A means to a final end is only wanted if an agent intends to bring about the final end, and hence also intends to bring about the means to this end. An agent may want to eat ice cream, but if the agent decides not to eat it, and therefore not to buy it, he or she does not intend to take the money out of his or her pocket. The agent does not want to do this, since there is no reason to do so.

Obviously, the above reasoning is based on the premise that means can only be determined by the use of reason. This assumption is justified because determining that something is a means to an end inevitably relies on reason. Furthermore, we should remember that our previous analysis showed that the distinction between desire and want relies on the fact that reason can be used by an agent to determine a want for something that is not determined by desire.

The conclusion of the above analysis is that the concept of intention, and more specifically the distinction between want and intention, is based on the fact that a want does not entail an intention to act according to the want. I shall argue in what follows (section 4) that this distinction can only be explained by the use of reason to evaluate our ends.

3. Reason, intention and the possibility of a multitude of ends

In order to demonstrate the distinction between want and intention I relied on an example in which an agent (desires and hence) *wants* to eat ice cream, but not *intends* to eat ice cream, due to dietary restrictions. It natural to interpret this example as a case in which an agent has more than one end.

Interestingly, previous attempts to refute the instrumental conception of practical reason have also relied upon situations in which an agent has more than one end, in order to support the claim that reason is used in a more substantial role than that of simply determining the means to an end (see for example, Hampton 1998, 167; Korsgaard 1998, 216–7).

Supporters of the instrumental conception of practical reason can, however, attempt to explain the distinction between want and intention by the possibility of an agent's having more than one desire, and hence more than one end. This can be explained as a situation in which one end, determined by desires, is put aside for the sake of another end, also determined by desires. This situation *prima facie* explains how it is possible for an agent to have a want (and a desire) for an end, while at the same time having no intention to pursue this end, thus explaining the distinction between want and intention, and hence the significance of the term "intention" itself.

Although I agree that the instrumental conception of practical reason fails to recognize that reason can be used to determine our ends, I do concede that the mere existence of a multitude of ends that an agent might have does not show that ends are susceptible to rational criticism. In what follows I explain why I believe that the mere multitude of ends that an agent might have ultimately fails to show that reason can be used by an agent to evaluate final ends, and hence fails to disprove the instrumental conception of practical reason.

To begin with, imagine a case in which there are two conflicting desires; for example, a desire to go to sleep and a desire to watch a film. The agent in this example wants both to go to sleep and to watch the movie, but cannot do both simultaneously. Hence, the agent must decide what to do. He prefers to watch the film, rather than going to sleep. He therefore watches the film, and later goes to sleep.

The previous scenario fails to explain the distinction between want and intention because it does not describe a case in which there is a want but no intention. The agent wants and intends to sleep (now), but will fulfil this intention only after the film ends. Notice also that according to this description of events (which for the moment I assume is coherent), reason is not used in order to intervene, and is not used by the agent to determine what he finally does. According to this description, the agent is pulled between two conflicting desires, and the stronger one prevails. Reason is not used in this scenario any more than in a case in which an agent is being pulled by two ropes in two different directions, and is forced by the pull of the ropes to move to the couch rather than to the bed.

According to the previous scenario, reason is not used to determine the action that the agent actually undertakes, which may make this example seem inappropriate for examining the use of reason in determining the intentions of an agent. However, its objective is to show that the distinction between want and intention cannot be explained by the multitude of ends per se, even if they conflict with one another, and that reason is not necessarily used in cases in which an agent has conflicting ends.

Let us now examine a second scenario, in which reason does intervene in

determining which action the agent actually undertakes. This example will show that conflicting desires *per se* cannot explain the distinction between want and intention.

Consider the first scenario again, but with the following alteration. The agent wants to sleep much more than he wants to watch the film. However, after considering all the relevant information, the agent decides to watch the film, and to go to sleep later. His consideration is that the film is shown only rarely, and he is not likely to fall asleep while watching the film.

In the second scenario, reason is, in fact, being used to determine the agent's action, and to decide between conflicting desires. However, even in this case, I can see how someone might argue that there is no gap between want and intention. The agent, who wants to sleep (now), still intends to sleep (now). However, it might be argued that his intention is withheld, and will only be fulfilled following watching the film.

The previous examples show that the mere multitude of ends that an agent might have, even if they conflict, fail to show that reason is nothing more than a "slave of the passions." This explains why previous attempts to disprove the instrumental conception of practical reason, based on situations in which an agent has more than one end, have failed. This is reflected by the failure of these scenarios to describe a case in which want and intention differ.

In order to describe a situation in which there is a want but no intention, we must think of a situation in which there are reasons for abandoning any attempt to pursue a specific end. Obviously, it cannot be presumed in the present context that reason can be used to evaluate a final end – that would beg the question. It is, however, possible to demonstrate the divergence between want and intention even in the context of instrumental reasoning.

Consider, for example, a situation in which conflicting ends force an agent to completely give up pursuing an end she desires. In this case, the agent has no intention to bring about this end. For example, assume that Naomi wants to go both to Ruth's party and Sara's party. She realizes the she can't go to both parties, since she has no way of getting from Ruth's house to Sara's house, or vice versa, in time to participate in both parties. Naomi decides to go to Ruth's party, and although she wants to go to Sara's party, she has no intention to do so.

Notice that in this scenario, reason plays an indispensable role in explaining the gap between want and intention. Naomi might fail to realize that she has to choose between the two parties, and decide to go to Ruth's party first (because she wants to go there more) and to Sara's party later. If she fails to recognize (using reason) that she has to forfeit going to Sara's party, she does not give up her intention to go to Sara's party. Hence, only reason can explain the disparity between her want and her

intention.

Since the previous example includes more than one end, and assuming that these ends are determined only by desire (any other assumption would obviously beg the question in the context of this discussion), it might be assumed to demonstrate that desire is responsible for explaining the disparity between want and intention. Thus, it may be admitted that the use of reason is *necessary* in order to explain the distinction between want and intention, since the recognition that two desires are in conflict, and that one should be discarded, is the conclusion of a reasoning process. However, it may further be claimed that the use of reason is *not sufficient* for explaining the divergence between want and intention, and that it is actually desire that cancels out the conflicting desire, thus explaining the possibility of a want without an intention.

I believe that this objection is the reason why previous attempts, which have relied merely on a multitude of (sometimes conflicting) final ends, have failed to refute the instrumental conception of practical reason. If this objection stands, it must be admitted that reason is not used in evaluating final ends, and is merely a 'slave of the passions.'

4. Reason as a critic of ends

So far I have argued that the mere possibility of an agent's having more than one end fails to disprove the instrumental conception of practical reason. In this section I turn my attention to an example in which an agent has only one desire, and only one end. In this case, supporters of the instrumental conception of practical reason cannot rely on conflicting desires in order to explain the distinction between want and intention. It is, rather, the use of reason itself that is responsible for rejecting an intention to bring about an end, which might be suggested by desire. Hence, intentional action relies on the use of reason by an agent in order to evaluate his or her suggested ends.

In order to acknowledge the use of reason as a critic of desires, imagine a case in which there is only one desire, that is, to go to Sara's party. Suppose that Naomi is too far from Sara's house to get to the party before it ends. Realizing that this is the case, Naomi gives up on going to the party. Obviously, she still wants to go, but she has no intention to go, because she knows that it is impossible for her to get there. Hence, participating in this event can no longer be described as her end.

This example shows us that ends are susceptible to rational criticism. An agent's ability to override the force of a specific desire by the use of reason does not depend on the force of another conflicting desire, but on the authority of reason. It is simply unreasonable to intend to realize an impossible end, and hence the use of reason enables an agent to reject a final end that might be suggested by desire.

In answer to this example, it might be objected that this example fails to show that final ends are susceptible to rational criticism, because it relies on an implausible identification of an end. It might be argued that the desire to go to Sara's party is more plausibly a means for enjoyment, for example. However, rather than supporting the instrumental conception of practical reason, this objection undermines this view. For it relies on a notion of "plausibility" that obviously implies a rational evaluation of potential ends, which has no place in the confines of an instrumental conception of practical reason. It is hence impossible for a supporter of this view to reject this example based on an objection to the hypothetical end.

The examples given above involve situations in which reason is used to *reject* final ends, in order to explain the conceptual distinction between want and intention. However, the conclusion that follows from analyzing these examples also has implications for situations in which reason is used to *approve* ends that might be suggested by desires. The connection between intentions, ends, and reason is a conceptual connection rather than a contingent connection, and it holds for every intentional action on the part of an agent. If reason is used as the critic of desires, and determines the intention to act, it also enables an agent to determine whether or not to satisfy a desire.

In order to demonstrate this point, imagine, for example, that Naomi wrongly determines that she is too far from Sara's house to get to the party before it ends. Naomi therefore gives up her intention to go to the party, and, hence, it is not her end. Suppose however that Naomi later recognizes her mistake. She can, in fact, get to the party in time. As soon as she realizes this, she decides to go to the party, and thus determines attending the party as her end. Notice that nothing has changed in her desires, and therefore nothing in her desires can explain why attending the party suddenly became a final end for her. The only change was in her recognition that it is possible for her to reach the party. In this example it is clearly the use of reason that enables Naomi to determine going to the party as her final end.

Final ends are therefore not determined by desire alone, but by the use of reason (perhaps together with a desire). The use of reason enables an agent to decide whether to satisfy desire. Therefore, although in these situations desire suggests certain ends, it is reason that enables an agent to determine that a certain desire should be satisfied, and it is the use of reason that *determines* a certain state of affairs as a final end for an agent.

It might be objected that a Humean subjectivist can accommodate this example. Surely, it might be argued, he or she need not deny that unsatisfiable desires fail to generate practical reasons. In the subjectivist's view, a practical reason is generated by a basic desire, plus facts relevant to its satisfaction. In cases in which an agent has a basic desire, but there is — whether by necessity or by happenstance — absolutely no means for

this desire's satisfaction (that is, no means to bring about its object), she has no practical reason to attempt to satisfy it. Hence, in forming no such intention, Naomi is showing her good sense, her virtuosity as a practical reasoner.

I believe that this objection misses the point of my argument. The fact that Naomi no longer intends to go to the party implies a change in her final end. As I have previously argued, an end for an agent is something that the agent intends to bring about. Therefore, if an agent does not intend to bring about something she wants, it is not an end for her. This is the reason why she has no practical reason to attempt to satisfy this desire. The object of this desire is not her end, and this can only be explained by the use of reason in order to evaluate final ends contrary to the instrumental conception of practical reason.

Another objection which might be raised at this point is that an instrumentalist may have an alternative explanation for the distinction between a want and an intention. Take the case of a single-desire agent who is irrational for not forming a corresponding intention. Why can't the instrumentalist simply say that an agent, when she does not form the intention to take the (presumably known) means to her only desired end, is simply means-end irrational? In such a case, this agent would seem to be guilty of a form of irrationality that (some versions of) instrumentalism can countenance. Accordingly, it is not necessary to appeal to a more "substantial" conception of reason to analyse this type of case.

However, this objection again undermines the instrumental conception of practical reason, rather than defends it. To begin with, it should be noted that it does not undermine the conclusion of the analysis of the previous example, according to which reason can be used in order to evaluate our ends. Furthermore, the claim that it is irrational of a single-desire agent to not form a corresponding intention implies that reason is used in order to determine intentions on the basis of wants – hence the failure is described as "irrationality" – and thus to determine the final ends of the agent, contrary to the instrumental conception of practical reason.

Finally, the previous examples focused on the possibility of *attaining* certain final ends, and may therefore give rise to the objection that they fail to show that reason can be used to evaluate and determine final ends. Although there is a sense in which it is justified to say that evaluating the possibility of achieving a certain end is an evaluation of this end, this seems to restrict the use of reason to its alleged instrumental function in determining means for ends. If, indeed, reason can be used to determine final ends, it could be expected that agents would be able to use reason to evaluate final ends as worth pursuing or not, in light of their intrinsic properties, rather than in light of the possibility of attaining these ends.

In response to this objection it is easy to show that reason can be used to

evaluate final ends as rational or irrational in light of their intrinsic properties. For example, suppose that Naomi decides to pursue her childhood dream of becoming a hairdresser and moving to a small village, where she will cut the hair of all those, and only those, inhabitants of the village who do not cut their own hair. She soon realizes this end is self-contradictory. A rational agent cannot adopt a self-contradictory scenario as his or her end, and realizing this, the agent decides to abandon the original intention to pursue his or her dream. Obviously, in this example it is also impossible to attain the end that the agent wants to pursue. However, the impossibility of attaining this end is due to the contradictory nature of the final end, which is recognized by the use of reason. The intrinsic properties of this end, rather than the mere nomological impossibility of attaining it, or a conflicting desire, therefore make it unreasonable as a final end.

Again, it might be objected that this example relies on an implausible identification of an end. However, as I have pointed out before, this objection undermines, rather than supports, the instrumental conception of practical reason. For it relies on a rational evaluation of potential ends, which has no place in the instrumental conception of practical reason. In fact, it is exactly my aim in this analysis – to show that ends are susceptible to rational evaluation.

In conclusion, the use of reason is a necessary and sufficient condition for explaining the distinction between want and intention. Reason is thus necessary for explaining the significance of the term "intention" and hence for explaining the possibility of an intentional action. It is the use of reason that enables an agent to determine his or her intentions, and hence her or her ends.

This conclusion explains why the existence of a desire for something (a) does not entail that the agent's desire is his or her end (e). Again, the mere number of desires, even if they are conflicting, cannot explain this distinction. A subject may not be aware of the fact that he or she has enough money, for example, to buy only ice cream or only chocolate. Only when the agent understands this does she dismiss the end of eating ice cream.

Hence, when used by an agent to determine his or her intentions, reason can be used either to approve or to reject his or her final ends, which might be suggested, rather than determined, by desires. According to the conclusion of the previous analysis, desires do not *determine* final ends, but only *suggest* final ends. If reason is used to determine intentions, it is also used to determine ends, which might be suggested by desires, which only then become an end for an agent.

5. Reason and the determination of ends

The previous section concluded with the understanding that it is reason that is used by agents to determine their final ends, rather than desire. Even if desire suggests these ends, it is ultimately the function of reason to approve or to reject these potential ends. My argument so far shows that reason can reject or approve ends. It might however be argued that this still does not give reason a sufficiently significant role. For a rationalist conception of practical reason implies that reason can be used to determine which ends *are worth pursuing* (and which are not).

In this section I will argue that reason can not only be used to evaluate final ends that might be suggested by desires, but *can* also determine final ends independently of desires altogether. I should stress that I do not merely argue that reason can determine ends that are not suggested by *actual* desires, but that reason can determine ends that are not suggested by anticipated desires. Rather, I argue that reason can determine final ends that are altogether independent of the agent's desire, whether actual or potential.

Defending this possibility is of great importance. Although the above analysis showed that reason can be used to determine final ends, the objection might be raised that it falls short of defending a rationalist conception of practical reason, as long as it is not shown that reason can be used to determine final ends that are not suggested by desires, whether actual or potential. Furthermore, if reason can be used in determining final ends, we might expect reason to also be used to determine final ends independently of an agent's desires.

A decisive reply to this concern is difficult to provide within the scope of this paper, since it requires giving an example in which reason can be used to determine a final end for an agent, without relying on any existing, or even possible, desire. This demand is equivalent to developing a full-blown rationalist theory of practical reason, which is beyond the scope of this paper. In the present context, I only attempt to defend the *possibility* of constructing such a theory. That is, I argue that reason *can* be used to determine final ends, but I do not argue that is actually does – a claim which obviously requires the identification of these ends. In what follows I explain how such a theory can be constructed.

To begin with, the previous example shows that it is possible to evaluate ends as rational or irrational independent of any desire. Naomi the hairdresser does not need to have any desire to fulfill her plan in order to realize that it is irrational to attempt to bring it about. All she needs to do is to consider this possibility.

Furthermore, suppose that reason can be used to determine that a rational agent should avoid certain states of affairs. This is not the case in the

previous example of Naomi's wish, because her end is self-contradictory, and therefore there is no need to attempt to avoid this alleged state of affairs. However, if reason can be used to determine that a rational agent should avoid certain possible states of affairs, it would follow that reason could *ipso facto* be used to determine which states of affairs (final ends) the agent should pursue.

It might be claimed that in this case reason is not used to determine a specific end, but rather to determine a *class* of states of affairs that the agent could rationally pursue. In answer to this argument it should first be noted that determining a final end always involves determining a *class* of states of affairs that the agent attempts to bring about. This is due to the fact that a state of affairs can be defined with the help of indefinite distinctions, and therefore can always be described more specifically or more generally. Thus, if an agent decides to eat ice cream, there is an indefinite number of events that would fulfill his or her intention. For example, he or she can eat vanilla or chocolate ice cream, from a cone or a dish, bought from an ice cream parlor or from an ice cream stand, and so on.

Furthermore, determining a state of affairs as an end for an agent is the same as determining a class of possible worlds that the agent should pursue; the agent should attempt to realize a member of the class. This is equivalent to determining that all other possible worlds, that is, those in which the agent does not eat ice cream, should be avoided (rather than brought about). The same end (for example, eating ice cream,) can be described negatively, as avoiding not eating ice cream.

Thus, determining which states of affairs should be avoided is equivalent to determining which states of affairs an agent should pursue. If it is possible to determine which states of affairs should be avoided, it is therefore also possible to determine which ones should be pursued.

An example of an attempt to develop and demonstrate this kind of rationalist theory of practical reason is found in Kant's moral theory (1996). Kant's theory is based on the categorical imperative, as a rational restriction on the ends and actions of any rational agent. This demands that the agent act only in accordance with that maxim that he or she can, at the same time, will to become a universal law. *Ipso facto*, it also supposedly determines a final end, that is, humanity, for any rational agent.

Obviously, I do not argue that Kant's theory is true. I do argue, however, that his strategy for developing a rational theory of practical reason is coherent. If this is correct, reason can be used to determine final ends, independently of desires altogether.

6. Conclusions

The starting point of my analysis was the use of reason by an agent in determining and evaluating the means for achieving certain ends, which might be suggested by an agent's desires. This is in accordance with the instrumental conception of practical reason, which denies that reason can be used in evaluating or determining final ends. However, contrary to the instrumental conception of practical reason, my analysis shows that reason can be used by an agent to evaluate, that is, to approve or reject, final ends that might be suggested by desires. This possibility is implied by the notion of intention, and hence is required in order to explain the nature of intentional action. Furthermore, my analysis shows that reason can be used to determine final ends independently of any desire, whether actual or potential.

The instrumental conception of practical reason is therefore inadequate, and should be replaced with a rationalist conception of practical reason, according to which reason alone can be used by an agent to determine which final ends he or she should pursue. Although I have not attempted to develop and defend a full-blown rationalist theory of practical reason, including principles of conduct that any rational agent should follow, I trust that the possibility of constructing such a theory has been sufficiently established.

REFERENCES

Andreou, C. 2005. The Voices of Reason. *American Philosophical Quarterly*. 42: 33–45.

Anscombe, G. E. M. 1957. *Intention*. Oxford: *Basil Blackwell.*

Audi, R. 1973. The Concept of Wanting. *Philosophical Studie*s 24: 1-21.

Bratman, M. 1984. Two Faces of Intention. *Philosophical Review* 93: 375–405.

Chisholm, R. 1964. The Descriptive Element in the Concept of Action. *Journal of Philosophy* 61: 613–625.

Davidson, D. 1963. Actions, Reasons, and Causes. *The Journal of Philosophy* 60: 685–700.

Hampton, J. E. 1995. Does Hume have an instrumental conception of practical reason? *Hume Studies* 21: 57–74.

Hampton, J. E. 1998. *The Authority of Reason*. Ed. R. Healey. Cambridge: Cambridge University Press.

Hubin, D. C. 1991. Irrational Desires. *Philosophical Studies* 62: 23–44.

Hubin, D. C. 2001. The Groundless Normativity of Instrumental Rationality. *The Journal of Philosophy* 98: 445–468.

Hume, D. 1978. *A Treatise of Human Nature*. Ed. L. A. Selby-Bigge. Oxford: Oxford University Press.

Kant, I. 1996. Critique of Practical Reason. In *Practical Philosophy*, ed. M. J. Gregor, 133–271. Cambridge: Cambridge University Press.

Korsgaard C. M. 1998. The Normativity of Instrumental Reason. In *Ethics and Practical Reason*, ed. G. Cullity and B. Gaut , 215–54. Oxford: Clarendon Press.

Lebar, M. 2004. Ends. *Social Theory and Pra*ctice 30: 507–533.

Markos, V. 2014. Instrumental Rationality. *European Journal of Philosophy* 22: 443–462.

Reitsma, R. L. 2013. How a Modern-day Hume Can Reject a Desire Categorically: A Perplexity and a Theoretically Modest Proposal. *European journal of analytic philosophy* 9: 48–66.

Schmidtz, D. 2001. Choosing ends. In *Varieties of practical reasoning,* ed. E. Millgram, 237–257. Cambridge, Ma.: The MIT Press.

Schueler, G. F. 1995. D*esire: Its Role in Practical Reason and the Explanation of Action*. Cambridge, MA.: MIT Press.

Searle, R. J. 1983. *Intentionality: An Essay in the Philosophy of Mind*. New York: Cambridge University Press.

Searle, R. J. 2001. *Rationality in Action*. Cambridge, MA.: MIT Press.

Smith, M. 1988. Reason and Desire. *Proceedings of the Aristotelian Society* 88: 243–258.

Spielthenner, G. 2012. Justificatory Reasons for Action. *European Journal of Analytic Philosophy* 8: 56–72.

Thompson, M. 2008. *Life and Action: Elementary Structures of Practice and Practical Though*t. Cambridge, MA.: Harvard University Press.

Van Roojen, M. 2002. Should Motivational Humeans be Humeans About Rationality? *Topoi* 21: 209–215.

Wasserman, R. 2011. Intentional Action and the Unintentional Fallacy. *Pacific Philosophical Quarterly* 92: 524–534.

Weller, C. 2013. Hume on the Normativity of Practical Reasons. *Hume Studies* 39: 3–35.

EuJAP | Vol. 13, No. 1, 2017
UDK
1 HUME, D.
165.81
165.172

"YES, THE THEORY IS ABSTEMIOUS, BUT . . .": A CRITIQUE OF YEHEZKEL*

REGAN LANCE REITSMA
King's College

ABSTRACT

This article is a critique of Gal Yehezkel's attempt to refute subjectivism about normative practical reasons, a school of thought inspired by Hume. Yehezkel believes reason, far from being, as Hume puts it, "the slave of the passions," has the normative authority to be a critic of basic desires and argues that subjectivism lacks the theoretical resources both to acknowledge this alleged truth and to analyze the distinction between wanting an outcome and intending to pursue it. I contend his refutation fails, largely because it operates with a strikingly attenuated view of the subjectivist theory.

Keywords: *Hume, practical reason, normativity, rationalism, subjectivism, realism, constructivism, instrumentalism, basic desires, personal ideals, maieutic ends*

> Reason is, and ought only to be, the slave of the passions, and can never pretend to any other office than to serve and obey them . . . 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an Indian or a person wholly unknown to me (Hume, 1888, 415-416).
>
> I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have *a problem*. Shall I act? Is this desire really a *reason* to act? (Korsgaard, 1994, 93, emphasis mine).
>
> Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action (Korsgaard, 1994, 91).

*Received: 23.10.2017.
Accepted: 02.11.2017.

1. Introduction

In "A Defense of a Rationalist Conception of Practical Reason," Gal Yehezkel tells us, "My aim is to show" Hume's "well-known" claim that reason "is, and ought only to be, the slave of the passions" is "false" (40). In Yehezkel's view, "advocates of the Hume-inspired subjectivist school of thought" are committed to an "instrumental" conception of practical reason that lacks the theoretical resources to account for reason's "role" within practical deliberation as a "critic of desire" (40, 45, 50-51). Moreover, he believes if we're to make the ubiquitously relevant—and very obviously real—distinction between merely wanting an outcome and intending to pursue it, we must adopt a notion of the faculty of practical reason "more substantial" than subjectivism can countenance (47).

Yehezkel is certainly going for the gusto. If his arguments were taken to be sound, this would shake the contemporary analytical landscape. Cullity and Gaut helpfully describe contemporary debates about practical reason as a perduring dispute between "three poles": Humean subjectivism, Kantian constructivism, and (Aristotelian) realism (1997); and within analytic philosophy, subjectivism is often considered the default position in theorizing about practical reasons (Nozick 1993, 133; see, also, Hubin, 1996; Millgram, 2001). Striking (at least for the purposes of this article) the pose of an agnostic between realism and Kantian constructivism, Yehezkel is convinced that, from his several arguments against subjectivism, "the true role of practical reason is revealed, and a rationalist view is established" (40)—"rationalism" is Yehezkel's name for the view that reason itself has the authority to "determine" whether any particular desire should be given a role in practical deliberation and what an agent ultimately ought to do.[1]

Yehezkel's argument is bold not only in philosophical import, but in argumentative tactics. His article is replete with the decisive terminology of proof, disproof, and deductive demonstration. His arguments "refute" subjectivism and its instrumental model of practical deliberation, and rationalism is "established" purely by means of conceptual analysis: his purported refutations do not, he tells us, appeal to "any contingent premises" (40, 51). Yehezkel wants to, intends to, take no philosophical prisoners: his opponent's view "should be replaced" (56).

Myself, I don't think Yehezkel's argument—whether taken as a demonstration, as a generic deductive argument, or as, say, an abductive

[1] Korsgaard gestures at one way to distinguish realist and constructivist conceptions of how, to use Yehezkel's terminology, rationalist-style reason "determines" how an agent ought to deliberate and behave (1994, 34-37). According to realism, the faculty of practical reason directly *discerns* objective normative truths; according to constructivism, the right answers to the relevant normative questions arise from applying the proper rational tests or procedures. Yehezkel's word, "determines," *seems* to be chosen because the word is ambiguous between "reason discerns" and "reason constructs."

argument—is sound. My basic contention is that Yehezkel's attempted refutation reflects, among other things, an uncharitable interpretation of the subjectivist school of thought and a crimped interpretation of how subjectivist-style practical reasoning is able to "serve" desire. There are tough questions for Humean subjectivists—I'll try to identify one at the end of this article—but I don't think Yehezkel himself has identified a weak point in subjectivist thinking.

2. What is Subjectivism?

Subjectivism about practical reason is, among other things, a theory about the ultimate source of normative practical reasons. Let's understand a normative practical reason, hereafter 'practical reason,' to be a consideration in favor of *doing* something.[2] It's a proper input into practical deliberation, a normative entity that ought to be factored into the process of deciding what to do.[3] Subjectivism claims that an agent's practical reasons, all of them, ultimately derive from among the "elements" of "his subjective motivational set, S" (Williams, 1993)—or, put more commonsensically, from among his own desires.

This claim—that practical reasons are ultimately grounded in basic, unmotivated conations—has been called, usefully, "the desire-based reasons thesis" (Hubin, 1999; Kagan 1992). To give a stock subjectivist example, your thirst—your basic, unmotivated desire for something to drink—gives you *a* practical reason to take the means to satisfy it, walking to the drinking fountain or to the juice bar, say. I will often put the subjectivist's defining idea in this way: in the stock example, your thirst is "rationally potent"; it generates practical reasons. The desire, along with facts relevant to its satisfaction, is the source of practical reasons for you.

Cut at its joints, subjectivism is committed, alongside the desire-based reasons thesis, to a second tenet, an instrumental principle that communicates normativity from a (rationally potent) basic desire to the (effective) means to its satisfaction. Together, these two elements make up the "core elements" of the subjectivist theory. Another characteristic of subjectivism—more difficult to define, but crucial to understanding this theory—is its *theoretical modesty*. Given its commitment to the

---

[2] I don't deny there are also considerations in favor of desiring, feeling, or being a certain way. But I will speak of practical reasons as considerations in favor of acting.

[3] I'll regard a practical reason as a *pro tanto* consideration: it does supply a positive consideration in favor of doing something, a consideration that ought to be weighed against whatever other practical reasons an agent happens to have, though it could conceivably be outweighed by one or more of them. For example, you might have a practical reason to take the afternoon off, but a weightier practical reason to put your nose to the grindstone.

desire-based reasons thesis, subjectivism regards practical reasons as *agent-relative*: the considerations relevant to answering the question "What, rationally speaking, should this particular agent do?" ultimately derive from among her own subjective, contingent, conative states—from among her own whims, impulses, desires, wants, cares, loves, intentions, pro-attitudes, and the like—and from nowhere else. Accordingly, subjectivism—in a noteworthy display of its quintessential *philosophical abstemiousness*—rejects any appeal to extra-subjectivist practical reasons or extra-subjectivist constraints upon practical reasons. That is—other than the instrumental principle itself[4]—a subjectivist cannot regard any purported normative standard, whether a realist standard of objective worth, a robust Kantian rule of practical reason such as the Categorical Imperative, or some other standard (such as the unnamed, unidentified standards, recognized by rationalist-style reason, implicit in Yehezkel's critique), as rationally binding upon an agent regardless of what she happens to want.[5]

This is to say that subjectivists believe, as Hume puts it, that practical reason's proper task—the only "office" to which it should "pretend"—is to "serve" ends the agent herself already desires. In a characteristically subjectivistic outlook, as Christoph Fehige puts it, "Some things are dear to our hearts. To act rationally . . . means in essence: to look after these things, as best we can" (2001, 49). Subjectivists often describe their view in the following way. When we ask an agent for a rational justification of his behavior, a chain of practical justifications properly "bottoms out" in an appeal to basic desires such as "because I was thirsty" or "because I care intrinsically about my child's welfare." In the subjectivist's view, practical rationality doesn't require that the agent have some further justification for why he should treat these basic desires as reason-giving. Motivational states such as these are the ultimate grounds, the fundamental starting points, of legitimate practical reasoning.

[4] If by 'categorical imperative' we mean a rule rationally binding upon an agent independent of the content of his contingent, subjective basic desires, the instrumental principle is such a principle (Dreier, 2001).

[5] Note, in a thorough-going subjectivist view (and following Hume, 1888), *even* the principle of prudence gains whatever rational validity it happens to have for a person ultimately from her own basic desires, either directly (from her intrinsic concern for her own [long-term] welfare), indirectly (from an instrumental need to take care of herself if she hopes to successfully pursue the satisfaction of her other [long-term] desired ends), or both (Hubin, 1979). The third and final sentence of the quotation from Hume above can be interpreted as reflecting this idea.

3. Skepticism about Subjectivism

As mentioned, subjectivism is often considered the default position in theorizing about normative practical reasons. It's generally touted as having a significant theoretical virtue. The agent-relative practical reasons it posits seem to have the *compelling force* we expect from practical reasons. Since according to subjectivism your practical reasons derive from among your own basic desires, you cannot "shrug off" these considerations; you can't properly say you aren't moved by the perspective from which they are generated (Hubin, 1996).

Put otherwise, when subjectivism levies at an agent the charge "You're behaving irrationally," the agent's grounds for taking heed are fairly clear. Imagine a very committed student, Desiree, who truly yearns for a good grade on a test, vividly knows she needs to study at length to get one, presently has no competing desires to do something else, and yet simply finds in herself no desire to crack open her book or to survey her class notes. She has a very weighty, even final practical reason to study, but there she sits, stultified. Desiree is guilty, it's natural to say, of "means-end irrationality." And since subjectivism treats practical reasoning as, at its very core, a "desire-governed" activity, this theory will be able to explain in a very intuitive way—to "any Desiree" it encounters—why it should matter *to her* that she has violated its standards. In violating subjectivism's standards, an agent is failing to "look after" her own desired ends.

All this said, subjectivism is a controversial theory. Obviously, it has its share of historically venerable opponents (many of whom I deeply admire). Hume's construction, "Reason is, and ought only to be, the slave of the passions" is a backwards-looking provocation. It's intended to subvert—rather boisterously—the long-standing idea, advanced by philosophers such as Plato, Aristotle, and Aquinas, that reason should rule. Disciples of such philosophers can and have wondered how these philosophers would best respond to Hume. Subjectivism's contemporary critics, whether on behalf of old philosophical visions or new, have raised many objections. Let's consider three.

First, several contemporary critics suppose subjectivism's theory-defining commitment to the rational potency of basic desires is subject to striking counterexamples: there are, this objection asserts, identifiable basic conations that simply don't generate practical reasons—people do, after all, find themselves wanting some very bizarre and some very awful things. Rachel Cohon, a realist, imagines a basic, unmotivated desire to stick one's finger in goo, finds it appropriate to judge this conation rationally impotent, and analyzes the case in this way: if the object of a basic conation lacks positive objective value, it is unable to generate practical reasons (2000, 63). Susan Wolf (2002)—who takes herself to be a more permissive, normatively-easier-going realist—judges that a desire

whose object lacks positive objective value but isn't objectively disvaluable (such as, perhaps, the desire to stick one's finger in goo) might generate practical reasons, but a desire whose object is objectively disvaluable (such as a vengeful impulse) does not.[6]

Second, another realist, Thomas Nagel, finds any Hume-inspired theory "glaringly incorrect" (2012, 106). After four decades of (impressive) philosophical investigation, there are, it is his much-scrutinized verdict, identifiable objective value judgments—such as 'it's wrong to torture animals'—that *must* be correct. He believes his intuitions about objective value, some of them intuitions about objective moral value, are sufficiently philosophically weighty to justify (i) rejecting subjectivism, (ii) developing a metaphysic that makes sense of these intuitions about objective value, and (iii) believing such objective values should carry strong weight within practical deliberation independently of any desires we happen (not) to have.

Third, some of subjectivism's contemporary critics (such as Christine Korsgaard) have objected—in very broad terms, and sometimes rather incredulously—that subjectivism is, in one or another way, an overly simple, even crude theory; to paraphrase the spirit of this complaint: "How can a theory with such a stripped down notion of the faculty of practical reason accurately represent the rich and complex processes of practical reasoning?"

Yehezkel's arguments are very much an expression of the third—very broad, incredulous—objection. Whereas Cohon and Wolf appeal to bizarre basic conations, Yehezkel believes that subjectivism fails to analyze any desires correctly—even very ordinary desires about going to a party and about becoming a hairdresser. As for a Nagel-style argument, Yehezkel very explicitly asserts he will not appeal to any substantive value judgments whatsoever (41, 54, 56). Instead, Yehezkel's aim is to show up the poverty of any instrumental conception of practical reason, and to do so purely through conceptual analysis, by reflecting on the ubiquitously relevant concepts of desiring, wanting an outcome, and intending to pursue an end.

So, how, we might ask, does Yehezkel *frame* his purported refutation?

---

[6] If you were to discover within your psychology an impulse to stick your finger in goo, a preliminary question would be whether this conation is basic or motivated. In an eleven year old boy of a certain disposition, it's likely to be a desire motivated by a more basic desire to gross out his friends—in which case the question whether his icky desire is a rationally impotent basic conation doesn't apply. That said, if such a conation did well up in me in the form of a basic conation, I'd be inclined to regard it—whatever its etiology—somewhat positively, say, as a small token of protest against today's prim standards of bourgeois cleanliness. (Perhaps there is in this seemingly small thought an important implication for subjectivism: namely, can't a person invest basic conations with new meaning—and with a role within her will—by "attaching" them to already held cares, concerns, and commitments?)

The following remarks are, I believe, helpful for thinking about Yehezkel's basic approach.

In the second (lengthy) quotation at the very beginning of this article, Korsgaard, a Kantian constructivist and "rationalist" in Yehezkel's sense, imagines what we might call the "ur moment of agency," the moment when a self-conscious being finds in herself a desire. As Korsgaard construes this moment, the agent asks herself, "Is this desire *really* a reason to act?" (1994, 93, emphasis mine).

Just to be clear, a subjectivist need not regard a basic desire "as" a practical reason: a basic desire is a mental state, a practical reason is a normative entity. Accordingly, I've described subjectivism as a theory that treats basic desires such as thirst and sleepiness as "sources" of practical reasons. That said, any subjectivist should take the slightly altered question, asked by an agent in the ur moment, "Should I treat this basic desire as rationally potent, that is, as the source of a practical reason for me?," to be a perfectly legitimate query. The simple truth is that a subjectivist will answer this question, at least in standard cases, 'yes'. (Does "in standard cases" sound weasely? Be patient; I'll address that objection, at length, in Section IV.)

Korsgaard makes it very clear she does *not* mean for this "Can I treat this desire as reason-giving?" question to be philosophically innocent. On the contrary, she's making a fundamental, reject-it-at-the-roots dismissal of the core subjectivist idea, namely, that practical reasoning "bottoms out" in appeals to basic, unmotivated desires. As she would have it, a basic desire, when initially recognized by an agent, presents not (as Hume himself would have it) a starting point for instrumental thinking, but a "problem." Korsgaard believes that any and every basic conation "must pass a kind of test of normativity before we can adopt it as a reason for acting" (1994, 91).

Similar to Korsgaard, Yehezkel's construal of the process of practical deliberation begins with the ur moment. Though he nicknames basic, unmotivated conations "suggestions" (42, 53, 54, 56), which is more friendly than Korsgaard's "problems," his basic framework is anti-subjectivist. A basic, unmotivated conation does not, by itself, generate any practical reasons whatsoever. Instead, every such conation must stand before the tribunal of practical reason and can be—in some cases, ought to be—"dismissed" (45) or "discarded" (50) by reason itself, which is presumably a form of categorical rejection that implies the relevant desire does not deserve to play any positive role in practical deliberation. As Yehezkel puts it, "reason can be used by an agent to evaluate, that is, to approve or reject, final ends, which might be suggested by desires" (40, 56).

A time or two, Yehezkel intimates that the philosophical cost of rejecting rationalism (and accepting the subjectivist's idea that all legitimate

practical reasoning is "governed by desire") is especially, even shockingly high. Take a "pure wanton"—my phrase, not Yehezkel's—to be a creature that, though it has the capacity for self-reflection, is "simply pushed and pulled by desires" (46) in a way analogous to a person "being pulled by two ropes in two different directions" (48). This creature never exercises, we're to suppose, any latent capacity to participate in practical deliberation and so persists as nothing more than an arena in which competing desires battle to be satisfied. Yehezkel seems to argue that if we are to believe that people are not pure wantons but are capable of forming intentions and so intending to act, we *must* adopt rationalism and its "more substantial" conception of reason (46). In his view, if rationalist-style reason does not "intervene" within practical deliberation —to turn desire's "suggestions" into rationally endorsed wanted outcomes, and rationally endorsed wanted outcomes into "pursuits"— then absolutely nothing can. Adopting subjectivism is, this is to say, tantamount to denying human agency.

4. Subjectivism and the Tribunal of Practical Reason

Note, Yehezkel's arguments presume that subjectivists are compelled by their theory to regard the faculty of practical reason as mute within the ur moment. When an agent finds in herself a basic desire, at "stage one," mum's the word from practical reason; it simply treats the basic desire as reason-giving and heads to "stage two" to do its proper task, seeking out means to its satisfaction. As I'll argue, this presumption, taken as a construal of modern-day subjectivism about practical reason, is mistaken. The core elements of subjectivism do not imply that the relevant "tribunal seat" is empty. Though a subjectivist will construe practical reason as incapable of a form of "intrinsic rational criticism" that realists and constructivists endorse, he can countenance various forms of "desire-governed" rational scrutiny of, and rational criticism of, basic conations. Let me explain.

No doubt, there is a reductionist strain within the subjectivist tradition. For example, some (famous) adherents of subjectivism seem to accept the view that all legitimate practical reasoning is instrumental in a very straightforward way (Russell, 1954). (This is the view, following other theorists, Yehezkel calls "pure instrumentalism" [40].) It will be relevant to recognize that the subjectivist school of thought as a whole is neither committed to pure instrumentalism, nor to a highly straightforward or simplistic model of practical reasoning.

Here is what I mean. Subjectivism certainly treats the process of instrumental reasoning—identifying basic desires; seeking effective means—as the *paradigm* activity of practical reasoning. For that matter, subjectivism also treats the process of weighing (sets of) competing

practical reasons against each other as a crucial step in determining the "final 'ought' before action." (More on that later.) When subjectivists provide illustrations of practical reasoning, they generally appeal to what we might call "garden-variety" examples of instrumental reasoning that seek out causal, criteriological, or mereological means to already-desired ends: being thirsty and seeking out behavior that will cause the thirst to go away, or wanting to run an officially-sponsored marathon and seeking out a race that meets the relevant criteria, or hoping to complete a particular twelve-step program and seeking out what the third step in the program actually is (Hubin, 1999).

There are, though, strains of subjectivism that posit forms of practical reasoning that are desire-based but don't fit the profile suggested by garden-variety examples. For instance, David Schmidtz (2001) argues that some legitimate practical reasoning is "maieutic": a matter not of finding right means to already-desired ends, but of seeking out and choosing new ends to desire. Say you find yourself wanting but lacking a sense of meaning in your life, and you come to the belief—let's assume justified and true—that this sense would come if only you were to have goals you care about intrinsically. You would thereby come to have a practical reason to seek out new goals—newly-desired ends—you can care about in this way. Since your choice of new basic desires is in the service of satisfying another basic desire (for a meaningful life), maieutic reasoning is agent-relative and desire-governed, and Schmidtz's proposal coheres, deeply, with subjectivism. The ultimate source of the practical advice 'choose a new end to desire' is the voice of one of your basic desires (for a meaningful life), enlightened by (accurate, we're assuming) reflections upon what it takes to satisfy it.[7]

Likewise, I myself have proposed a structural complexity that is, against the historical grain, available to subjectivism: a form of "categorical" reasoning, the disenfranchising—or "dismissing" or "discarding" or "silencing"—of basic desires that violate a cared-about personal ideal (Reitsma, 2013). As I see things, in the run of life we often do treat our basic, unmotivated desires as reason-giving: you feel thirsty and you take yourself to have a reason to drink; you get sleepy and take yourself to have a reason to sleep. Occasionally, though, we find in ourselves an impulse or desire that gives us pause, or even horrifies us. In such a case, a person might proclaim, "I can't treat *that impulse* as reason-giving," or —with an interestingly different inflection—"*I* can't treat that impulse as reason-giving." One question is whether a subjectivist can, in good standing with the core elements of her theory, interpret *some* such cases as instances in which a basic desire is rendered rationally impotent. I've

[7] Yehezkel says, "By an 'end' I simply mean a final end, that is, something we pursue for its own sake—rather than . . . a maieutic end" (43). But this confuses the shape of a maieutic end with its adoption. A maieutic end is adopted so as to satisfy another desired end; but once adopted, its object is itself desired, and so pursued, for its own sake.

argued that she can.

Here's one way how. My proposal appeals to the idea that at least some people care about what we might sensibly call "personal ideals." I regard a personal ideal as a normative standard that generates substantive practical advice. Some personal ideals are very complex. A personal ideal might answer a wide array of practical questions about how (not) to behave, how (not) to weigh practical reasons (against each other), how (not) to feel, and what character traits (not) to foster. Consider, for instance, the case of a loving parent. At the heart of a loving parent's love is her desire for her child's present and long-term welfare. Let's imagine that a particular loving mother not only cares about her children, but also strongly desires to *be* a good parent. This mother's ruling passion and her partially corresponding personal ideal make demands upon her. Behavioral demands: she ought to feed and clothe her child. Emotional demands: she ought to experience characteristic patterns of emotional concern for her children. And volitional demands: she ought to treat the fact that a course of action will significantly benefit her child as, in the very least, a significant practical reason to do it. A ruling passion might also call for, in the run of a particular devotee's life, the adoption of personal guidelines. For instance, a loving parent who consistently loses patience with her children, or who sometimes finds herself resentful of her children's "neediness," might decide to adopt a maxim "Be more patient" or "Pay more attention to the good things that come from having children." As I see it, these guidelines are properly incorporated into her own practical point of view, into her will. The voices of her love and her desire to be a good parent deserve a privileged seat at the table, with the practical authority—in some cases—to demand that the agent make significant changes in behavior, volition, and character and that the agent adopt maxims to help her pursue one of her ultimate goals, helping her children flourish.

Here is the crux of the matter, at least with respect to my proposal. Among its various volitional demands, a personal ideal might place restrictions on what a devotee can treat as reason-giving. Borrowing an example from Gary Watson (1982), imagine that a devoted mother finds in herself, rather out of the blue, an utterly uncharacteristic desire to drown her beloved infant. A good parent will not treat, this mother sensibly supposes, a violent impulse against her child as generating a reason. This impulse is not merely 'trumped', or even 'swamped', by the weightier "love-based" practical reasons she has. This impulse is, for the parent, rationally impotent and so deserves to be silenced."[8] The mother's personal ideal includes, we might say, "norms of rational impotence" that

[8] Note, this tweaks the meaning of the Aristotelian John McDowell's word 'silenced' (1979). For McDowell, a "silenced" consideration doesn't even arise in the virtuous agent's practical deliberation. In this mother's intended usage, to say a desire is "silenced" is to say it is treated as categorically unworthy of consideration.

demand this response. Since in my proposal it is the agent's caring about the ideal that makes the ideal and whatever norms it includes normative for her, my proposal also posits a desire-governed but non-garden-variety form of practical reasoning and so coheres, deeply, with the core elements of subjectivism.[9]

These two forms of practical reasoning are more complex in structure than a subjectivist model (such as pure instrumentalism) that includes only varieties of instrumental reasoning akin to the "garden-variety" cases. In both models, reason—desire-governed reason—plays a "more substantial" role in practical deliberation than seeking causal, criteriological, or mereological means. In the model presented by Schmidtz, desire-governed reason is playing a "role" in advising the agent to adopt, and so—if reason's practical advice is successfully heeded—in generating, a newly desired end. In my own proposal, desire-governed reason is involved in appraising—and sometimes in "discarding" or "dismissing"—a basic conation. In both of these cases, though, the "shape" of reason is not "rationalist."

Perhaps, to stave off my reader's incredulity, it's important to point out that the desire-based reasons thesis, though it does state that *all* practical reasons are grounded in basic desires, does not imply, in and of itself, that *every* basic desire generates practical reasons. If we ask the question, "Is it possible for a subjectivist to grant that there are rationally impotent basic conations?," the desire-based reasons thesis does not logically rule out a yes answer. My proposal not only recognizes this logical possibility, it presents a positive case for the claim that some basic desires are rendered, by desire-governed reason, rationally impotent.

There are at least two other ways "desire-governed reason" might occupy the seat of a tribunal that rationally scrutinizes basic conations. First, it is possible for an agent to come to have an instrumental reason to (strive to) eradicate a "mere want." That is, desire-governed reasoning can sensibly lead an agent to consider whether merely having a particular desire is detrimental to the successful pursuit of other strongly desired ends. A father desperately fighting an addiction, partly so that he can live up to his desire to be a good father, will likely recognize he has an extremely strong instrumental reason, if he can discover the means, to eradicate his addictive impulses. Second, some basic desires are *unsatisfiable*, and so, given that there are no means to their satisfaction, do not generate any practical reasons. In some such cases, it is a contingent fact that there are no means to the relevant desire's satisfaction: a person craves the last donut in the box, but his hands are chock full of books such that it's painfully obvious he can't pick up the donut before someone else snags it. Also, conceivably a person could find himself wanting a logically impossible end, akin to desiring to have your cake and eat it, too' or

[9] For a critical discussion of my proposal, see Rippon (2014).

desiring that '2 + 7 = 11'. A subjectivist can, consistent with her theory, regard unsatisfiable desires as unable to generate practical reasons.

To summarize and to suggest some useful terminology, a subjectivist might argue for the "default" rational potency of basic conations. Let's describe things in this way: whereas Korsgaard favors treating a basic conation as in and of itself "rationally impotent" and needing to "earn," through rationalist-style reason's endorsement, the "normative right" to play a positive role in practical deliberation, the subjectivist is committed, in the very least, to the idea that a basic conation is "rationally potent unless rendered otherwise." In other words, subjectivists will generally presume that a basic conation is rationally potent, but they are able to employ models such as Schmidtz's or mine or the two mentioned in the previous paragraph to handle non-standard cases. The general point is that there are ways for a subjectivist to account for "reason's role" in generating newly desired ends and "reason's role" in rationally scrutinizing basic desires. We'll have to see whether these theoretical resources are sufficient to address Yehezkel's purported refutation, his consistent refrain that, unless we appeal to "rationalism," it will be impossible to analyze cases of intending to act, or wanting something without intending to pursue it.

5. Yehezkel's Main Argument

So, what is the structure of Yehezkel's purported refutation?

The author distinguishes between desiring, wanting, intending, and acting intentionally (42-47). The distinction most clearly central to the author's argument is between wanting [an outcome] and intending [to bring about that outcome] (46-47). This distinction, as I said earlier, is very obviously real. Following one of Yehezkel's examples, Penny might want to eat ice cream, but have no intention to devour any, perhaps because she has firmly resolved to save money. Also, Naomi might want to attend Susan's party, but—since she realizes the party is almost finished and she's miles and miles away—not form any intention whatsoever to travel in its direction.

The crucial question, Yehezkel believes, is "How best to analyze such examples?" In Yehezkel's view, the best—the correct—model of practical reasoning is rationalism: a rationalist form of reason, he thinks, plays a role in every decision whether to treat a basic conation as reason-generating and every decision about whether to intend to do what is wanted. But Yehezkel can't simply *assert* that his anti-subjectivist theory of practical deliberation is correct; that would beg the question. This prompts him to search for a "proof" that subjectivism is incapable of analyzing examples—in his view, any examples—of wanting an outcome without intending to pursue it.

Yehezkel doesn't think this—the discovery of a proof—has been accomplished within the philosophical literature. Here's his diagnosis:

> previous attempts to refute the instrumental conception of practical reason have . . . relied upon situations in which an agent has more than one end, in order to support the claim that reason is used in a more substantial role than that of simply determining the means to an end . . . Supporters of the instrumental conception of practical reason can, however, attempt to explain the distinction between want and intention by the possibility of an agent's having more than one desire, and hence more than one end . . . one end, determined by desires, is put aside for the sake of another end, also determined by desires. This situation *prima facie* explains how it is possible for an agent to have a want . . . for an end, while at the same time having no intention to pursue this end (47-48).

In other words, the subjectivist can always say, about any situation in which a person has competing desires, "Well, Penny wants to eat ice cream but doesn't intend to eat any because there is something else, inconsistent with eating ice cream, that she wants even more." That is, the subjectivist will treat Penny's "firm resolution" to save money as grounded in a desire that is presently in competition for her hankering for something sweet.

In response, Yehezkel makes this judgment: if there is, in a particular case, an explanation of how the agent has a want to X without a corresponding intention to X, an explanation that is consistent with 'instrumentalism', then 'instrumentalism' (though false) isn't yet refuted. Accordingly, Yehezkel believes anti-subjectivists should construct an utterly new type of example: a case of an agent with only one desire. This type of case is best able to *reveal*—to conclusively prove—that we must appeal to rationalist-style reason to countenance intending to act (50-55). This appeal to a new type of example is what makes his argument "novel" (40, 49-50).

Structurally, Yehezkel argues in what we might call a "reason of the gaps" strategy: if, to analyze a particular case of wanting but not intending to act (especially of a single-desire agent wanting but not intending to act), we appeal to no more than the concepts (basic desires, means-end reasoning) that any and every subjectivist conception of practical reason permits, then we are left at an impasse and we thereby discover that "reason can intervene" (47) or "reason must enable" (46) or we "can only turn to reason" (45) to equip us to analyze the act.

Here is Yehezkel's allegedly debate-clinching example:

> Suppose that [single-desire] Naomi is too far from Sara's house to get to the party before it ends. Realizing that this is the case, Naomi gives up on going to the party. Obviously, she still wants to

> go, but she has no intention to go, because she knows that it is impossible for her to get there. Hence, participating in this event can no longer be described as her end (50).

The striking conclusion:

> This example shows us that [desired] ends are susceptible to rational criticism. An agent's ability to override the force of a specific desire by the use of reason does not depend on the force of another conflicting desire, but on the authority of reason. It is simply unreasonable to intend to realize an impossible end, and hence the use of reason enables an agent to reject a final end that is suggested by desire (50).

What to make of this argument?

6. A Subjectivist's Response

The target of this attempted refutation, the subjectivist, won't be very impressed, whether he is inclined to accept a simpler or a more complex strain of the Humean school of thought. I suppose some subjectivists might harbor a suspicion that the example of Naomi presumes other basic desires: if Naomi is a rather ordinary person, she presumably doesn't want to expend physical energy, or significant gas money, on a foolhardy, wild goose chase. But Yehezkel would likely accuse such a move of begging the question: he has stipulated, after all, that the desire to go to the party is Naomi's only desire.

It's not clear whether Yehezkel is asking us to conceive of Naomi as a quite unusual creature, a philosopher's fancy, who literally has one and only one basic desire for all or most of her life. (If someone dropped an anvil on Naomi's foot, would her desire to go to Sara's party remain her sole conation, her only fixation?) Or, alternatively, whether Yehezkel would like us to see Naomi as a "more ordinary" person who, caught in a brief episode of one-track-mindedness, has one and only one desire present in her soul at the moment. I'm not sure it matters, though. As I see it, however we interpret Naomi's psyche, there are other, more serious problems for Yehezkel's argument.

First, as discussed earlier, a subjectivist simply needn't deny that *unsatisfiable* basic desires fail to generate practical reasons. In the subjectivist's view, a practical reason is generated by a basic desire, plus facts relevant to its satisfaction. In cases in which there is, even by simple happenstance, no means to the desire's satisfaction (i.e. its object obtaining), there is no practical reason. This is to say that a Humean subjectivist would, same as Yehezkel, regard Naomi as lacking any practical reason to attempt to travel to the party. The proximate dialectical upshot: subjectivism has the theoretical resources to analyze

this purported counterexample.

Second, stepping back into the broader dialect, Yehezkel's would-be clincher simply doesn't strike me as the sort of example that might knock a tottering subjectivist (such as myself, by the way) into the arms of anti-subjectivism. Merely wanting to go to a party isn't the kind of desire realists or Kant, taken as a constructivist, typically find "glaringly" contrary to reason.

Myself, I have ears to hear the call of a Nagel-style argument against subjectivism's implicit rejection of the objectivity of value. Nagel believes some intuitions about the objectivity of moral value should be treated as having profound metaphysical import. Though I'm not inclined to adopt Nagel's particular metaphysic, when I see, for example, a vulnerable human being treated with contempt, I feel the force of Nagel's desire to build a metaphysic that makes "deep sense" of the idea that such contempt is blind to the reality of person's true value.

Also, given the odd quirks of human psychology, and the shocking, knavish, and cruel loves of seemingly morally pernicious people (imaginary or real), I can feel the strong pull of Cohon- and Wolf-style arguments that appeal to realist norms. Truly suicidal thoughts experienced by an otherwise well-off person, petty and vengeful desires, brute contempt for the weak, hatred or disdain for sentient animals, these are basic conations that might intuitively be charged with being intrinsically demerited. Does Naomi's desire deserve anything like the same sort of criticism?

I don't think so. To my thinking, Naomi's desire *itself* does not warrant the type of categorical rational criticism—"dismissing" or "discarding"—Yehezkel seems to levy at it. We can see, I submit, why someone who cares about behaving, and so deliberating, like a good parent would find a basic desire to harm her child as itself objectionable. But if, on Yehezkel's conception, rationalist-style reason judges Naomi's happy-go-lucky desire intrinsically irrational, reason might come to seem—not so much a helpmate for the agent's heart or for his deepest moral convictions, but—a dour and seriously unfun faculty.

I'll press a little harder. Doesn't a subjectivist analysis of this case have, at least at a glance, notable *advantages* over Yehezkel's? Subjectivism takes our natural tendency to treat our basic desires as reason-giving and endorses the general run of this rampantly ubiquitous practice—as I've been saying, subjectivism has an "easy way about it" that complex strains of subjectivism attempt to nuance. Wouldn't a sensible theory of practical reason likewise grant Naomi's desire to go to the party, even if it turns out, by happenstance, to be impossible to satisfy, *some* measure of normative weight? In the present circumstances, this basic desire doesn't generate a practical reason for Naomi to, say, get in a car. But the desire

would seem to justify other attitudes she might have. What if circumstances changed? If by chance Sara decided to change her party to a more favorable date, Naomi's desire to attend, once there is a means to its satisfaction, would presumably generate a practical reason. Accordingly, what if Naomi began to *hope* Sara will choose to change the party's date to another day? Wouldn't Naomi's wanting to go to the party play a role in rationally justifying this hope? Moreover, think of Naomi's feelings. Let's say Naomi, realizing she can't make it to Sara's party, indulged in a few moments of "aw, darn" consternation that she has to miss the party; it would help to rationally justify her emotional response if we were to see the desire as characteristically capable of generating practical reasons. The subjectivist can say Naomi's mild displeasure is warranted, since she's missing out on something that (to one degree or another) matters to her. The point: this rather innocuous desire doesn't seem to be the kind of desire that ought to come in for robust rational criticism. Unlike the mother's out-of-the-blue violent impulse, which is condemned by the mother's cared-about personal ideal, Naomi's doesn't warrant being "targeted" by reason.

7. Summing Up

I've argued that, though a subjectivist is committed to denying that a satisfiable basic conation can be *intrinsically* rationally impotent, she is able to conceptualize the faculty of practical reason such that it plays the role of a "critic" of basic conations. The theory is abstemious, but not nearly as limited as Yehezkel supposes. I find it important to say, too, that Yehezkel's claim that a subjectivist theory cannot make sense of the notion of an intentional action seems extravagant to me: is, for instance, the Bratman-style (1999) idea that intentions are partial plans unavailable to a subjectivist?

My critique could, in some reader's minds, raise the question whether I have simply, flat-out misunderstood the structure of Yehezkel's argument. I've wondered this myself. (If I have misconstrued his arguments, may Yehezkel live up to his name and find it in him to forgive me.)

However, Yehezkel's arguments seem to me, in crucial ways, underdeveloped and vague. To argue that a school of thought lacks the theoretical resources to analyze important concepts or important examples would *seem* to require the critic to display this school of thought at its very best and subsequently point out that the theory, even in tip-top form, is unable to account for the relevant ideas. Yehezkel, however, appears to direct his objections at the most minimal form of subjectivism—pure instrumentalism—not at its most sophisticated. Yehezkel himself says that pure instrumentalism "is the starting point of my analysis" (40), and at no point does he critically engage a less

minimalist subjectivist account and its theoretical resources.[10] Also, instead of providing definitions, or in the very least helpful glosses, of instrumentalism, subjectivism and rationalism, he tends to employ stock phrases.[11]

Moreover, my interpretation of the structure and content of Yehezkel's reasoning gains credibility if we consider the other bold argument he makes. Recall, Yehezkel argues that if rationalist-style reason doesn't "intervene" to endorse or reject a particular basic desire, there's simply nothing else that can, and the relevant person will be at the mercy of whichever conation happens to be the causally most powerful. In such an event, this person is not, Yehezkel intimates, truly an agent at all, but a "pure wanton," a creature merely "pushed and pulled" by desires (46; see also, 48). But doesn't this argument simply conflate subjectivism and behaviorism? If Yehezkel's conception of subjectivist-style reason is one according to which "reason" does nothing more than stand by as desires simply overpower the creature, it is an attenuated conception of subjectivism, indeed!

Otherwise put, when a Humean subjectivist regards reason as "the slave to the passions," she needn't suppose the faculty of practical reason is, as it were, gagged and straightjacketed, unable to participate in the causal processes between discerning desires and behavior. Instead, the sense in which practical reason is, in the subjectivist view, "the slave," is that the faculty of practical reason takes its "bidding," it's substantive orders and practical advice, from what the agent, as Fehige put it, "finds dear" (2001, 49). And, as I've tried to add, some desired ends are sufficiently complex that they make demands on how an agent ought to treat the conations that happen to well up in her day-to-day psychology.

---

[10] Yehezkel does mention subjectivists who reject pure instrumentalism (40), but he doesn't, within what I've called his "main argument," attempt to reveal that such subjectivists—theorists who seemingly have developed more theoretical resources than the pure instrumentalist—can't analyze the cases he constructs.

[11] I see a penchant within Yehezkel's article to use the word 'reason' in a way that's unhelpfully vague. In my view, Yehezkel should very clearly distinguish between rationalist-style reasoning and desire-governed reasoning and construct examples in which it "must" be rationalist-style reason, *not* a form of desire-governed reason, that is doing the relevant work within the agent's practical deliberation. In several cases, however, the "use of reason" Yehezkel identifies is, in the very least arguably, an instance of garden-variety instrumental reasoning. In such cases, I'm inclined to agree—"yes, reason is doing the work"—but, contrary to the thrust of Yehezkel's argument, it's nothing more than instrumental reason.

8. Where To From Here?

Despite my criticisms of Yehezkel's arguments, there are, of course, tough questions for subjectivists. There always are for theories that cut to the philosophical bone. Let me attempt to broach one particular hard question, perhaps suggested by Yehezkel's arguments, for "complex" strains of subjectivism that appeal to the existence of such philosophical constructs as "cared-about ideals" and "ruling passions."

Subjectivists very often distinguish between an agent's cares and her "mere wants." I myself have consistently done this in the stock subjectivist examples I've provided. For instance, I've treated the mother's love for her child as a "ruling passion" and her desire to be a good parent as a "ruling passion" that is also a "cared-about ideal." Whatever else a "ruling passion" is, it's supposed to be a conation that, in the complex subjectivist's thinking, deserves special authority within the relevant agent's practical reasoning: for instance, in normal circumstances, when a ruling passion is in conflict with a "mere desire," the ruling passion generates weightier practical reasons that ought, ultimately, to be acted upon.

"How," a critic (or, for that matter, a thoughtful, self-reflective subjectivist) might ask, "do some basic conations gain, for a particular agent, more normative authority than others?" Calling a basic desire a "ruling passion" obviously isn't enough. The phrases 'mere desire' and 'ruling passion' mark the distinction, but they don't answer the question.

Some conceivable answers certainly won't do. Is it merely that the parental desire is more intense, in a phenomenological sense, than the violent whim? That doesn't seem right: if the whim increases in intensity, does it threaten to become a ruling passion? The sorts of subjectivists I'm thinking about—"complex subjectivists" who think practical reasoning is about "looking after" what one finds "dear"—certainly won't think so. Is it merely that the parental desire has greater causal power? But then we don't seem to be talking so much about which desire *deserves* greater normative authority; we're simply heading in a behaviorist direction, according to which the "right" action is whatever behavior happens to occur.

"So," the question for the "complex subjectivist" is, "what *is* your account?"

Let's make a distinction between two examples of the ur moment: an "in-the-midst-of-life" ur moment and an "earlier" such moment.

Within the philosophical literature, there are much-developed accounts of how ruling passions differ from mere desires, and why—in the midst of an ordinary, mature person's life—her already developed "loves" and "cares" deserve special normative authority. Harry Frankfurt (1988, 1999,

2004, 2006), for instance, can be seen as having spent the better part of a distinguished career developing this type of account. The loving mother who feels a violent impulse to harm her child, it is worth noting, confronts this particular "ur moment" with an already well-developed practical point of view. She knows what she finds dear. And perhaps the complex subjectivist is able to account for why her particular cares and concerns should carry so much weight for her. When the mother follows the dictates of her ruling passions, we can see her behavior as rational in the very sense stultified Desiree's isn't.

But consider a developmental version of the above question. You and I haven't always had a well-developed conception of what we care about. Once, at whatever age we were, we confronted desires without yet having a strong sense of what kind of person we desired to be. Noting this might lead to the question: "How, dear subjectivist, from the 'raw material' of the conations a developing person—a budding agent—happens to find within herself, plus practical thinking, do ruling passions arise? When an agent without ruling passions confronts his basic conations, how does he properly decide what to do or what, more significantly, he cares about?"[12] Accordingly, a version of the earlier anti-subjectivist suspicion might naturally arise: "Reason, a more robust type of reason than subjectivism countenances, must play some role in identifying, within a growing person's thinking, which basic conations *deserve* to be granted a high level of practical sway within the agent's life." Did the loving mother, after all, simply "opt" in some "brute way" to prefer her affectionate impulses to whatever other desires she happened to find in herself?

I suspect many skeptics of subjectivism will suppose the development of a heart, the development of a personal or practical point of view, a perspective from which some things come to be "dear" to a person, requires a type of practical guidance that a Humean will be hard-pressed to countenance.

REFERENCES

Bratman, M. 1999. Intentions, Plans, and Practical Reason. Cambridge: Cambridge University Press.

Cohon, R. 2000. The Roots of Reasons. *The Philosophical Review* 109 (1): 63-85.

[12] For a similar challenge to Frankfurt's "complex" subjectivist theory of practical reason, see Reitsma (2013, 62-64).

Cullity, G. and B. Gaut, eds. 1997. *Ethics and Practical Reason*. Oxford: Oxford University Press.

Dreier, J. 2001. Humean Doubts about Categorical Imperatives. In *Varieties of Practical Reasoning*, ed. E. Millgram, 29-47. Cambridge, MA: MIT Press.

Fehige, C. 2001. Instrumentalism. In *Varieties of Practical Reasoning*, ed. E. Millgram, 49-76. Cambridge, MA: MIT Press.

Frankfurt, H. 1988. *The Importance of What We Care About*. Cambridge: Cambridge University Press.

_____. 1999. *Necessity, Volition, and Love*. Cambridge: Cambridge University Press.

_____. 2004. *The Reasons of Love*. Princeton, NJ: Princeton University Press.

_____. 2006. *The Tanner Lectures in Moral Philosophy*, ed. D. Satz. Stanford, CA: Stanford University Press.

Hubin, D. 1979. Prudential Reasons. *Canadian Journal of Philosophy* 10: 63-81.

_____. 1991. Irrational Desires. *Philosophical Studies* 62: 23-44.

_____. 1999. What's So Special about Humeanism? *Noûs* 33 (1): 30-45.

_____. 2003. Desires, Whims and Values. *The Journal of Ethics* 7: 315-335.

Hume, D. 1975. *An Enquiry Concerning the Principles of Morals*. 1888., ed. L.A. Selby-Bigge. 3rd edition. Oxford: Clarendon Press.

_____. 1978. *A Treatise of Human Nature*. 1748., ed. L.A. Selby-Bigge and P.H. Nidditch. Oxford: Oxford University Press.

Kagan, S. 1992. The Structure of Normative Ethics. *Philosophical Perspectives* 6: 223-242.

Korsgaard, C. 1994. *The Sources of Normativity*. Cambridge: Cambridge University Press.

McDowell, J. 1979. Virtue and Reason. *Monist*: 331-350.

Millgram, E. 2001. Practical Reasoning: the Current State of Play. In *Varieties of Practical Reasoning*, ed. E. Millgram, 1-26. Cambridge, MA: MIT Press.

Nagel, T. 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature is Almost Certainly False*. Oxford: Oxford University Press.

Nozick, R.1973. *The Nature of Rationality*. Princeton, NJ: Princeton University Press.

Reitsma, R. L. 2013. How a Modern-Day Hume Can Reject a Desire Categorically: A Perplexity and a Theoretically Modest Proposal. *European Journal of Analytic Philosophy*. 9: 48-66.

Rippon, S. 2014. On the Rational Impotence of Urges. *European Journal of Analytic Philosophy*. 10: 70-75.

Russell, B. 1954. *Human Society in Ethics and Politics.* London: George Allen and Unwin.

Schmidtz, D. 2001. Choosing Ends. In *Varieties of Practical Reasoning*, ed. E. Millgram, 237-256. Cambridge, MA: MIT Press.

Smith, M. 1987. The Humean Theory of Motivation. *Mind* 56: 36-61.

_____. 1994. *The Moral Problem.* Oxford: Blackwell.

Watson, G. 1982. Free Agency. In *Free Will*, ed. G. Watson, 96-110. Oxford: Oxford University Press.

Williams, B. 1993. *Moral Luck*. Cambridge: Cambridge University Press.

Wolf, S. 2002. The True, the Good, and the Lovable: Frankfurt's Avoidance of Objectivity. In *The Contours of Agency*, ed. S. Buss and L. Overton, 227-244. Cambridge, MA: MIT Press.

FISSION, FIRST PERSON THOUGHT, AND SUBJECT-BODY DUALISM

FISIJA, MISAO IZ PRVOG LICA I DUALIZAM UMA I TIJELA

KIRK LUDWIG
Indiana University

ABSTRACT

In "The Argument for Subject Body Dualism from Transtemporal Identity Defended" (2013), Martine Nida-Rümelin (NR) responded to my (Ludwig 2013) criticism of her (2010) argument for subject-body dualism. The crucial premise of her (2010) argument was that there is a factual difference between the claims that in a fission case the original person is identical with one of the successors. I argued that, on the three most plausible interpretations of 'factual difference', the argument fails. NR responds that I missed the intended, fourth interpretation, and that the argument on the third interpretation goes through with an additional assumption. I argue that the fourth interpretation, while insufficient as stated, reveals an assumption that provides an argument independently of considerations involving fission cases: in first person thought about future properties we have a positive conception of the self that rules out having empirical criteria of transtemporal identity. However, I argue that the considerations offered for this thesis fail to establish it, and that we do not, in fact, bring ourselves under any positive conception in first person thought, but rather think about ourselves directly and without conceptual mediation. This explains why it appears open in fission cases that the original person is identical with one of the successors, while what is possible is constrained by the actual nature of the self as referred to in first person thought. I argue also, incidentally, that on the third interpretation, the first premise of the argument is inconsistent with the necessity of identity.

Keywords: subject-body dualism, fission cases, first-person thought, transtemporal identity, Martine Nida-Rümelin

SAŽETAK

U *The Argument for Subject Body Dualism from Transtemporal Identity Defended* (2013), Martine Nida-Rümelin (NR) odgovara na moje kritike (Ludwig 2013) njenoga argumenta (2010) za dualizam subjekt-tijelo.

Ključna premisa njenoga argumenta (2010) jest da postoje faktične razlike između tvrdnji da je u slučaju fisije, izvorna osoba identična s jednim od nasljednika. Ja argumentiram da, na temelju tri najplauzibilnije interpretacije *faktične razlike*, argument pada. NR odgovara da nisam uočio namjeravanu, četvrtu interpretaciju, i da se argument treće interpretacije provlači s dodatnom pretpostavkom. Odgovaram da, dok je četvrta interpretacija nepotpuna u načinu na koji je formulirana, ona otkriva pretpostavku što pruža argument neovisno o razmatranju koje uključuje slučajeve fisije: u misli prvog lica o budućim svojstvima, imamo pozitivnu koncepciju jastva koje isključuje empirijski kriterij za transtemporalni identitet. Međutim, argumentiram da razmatranja ponuđena u prilog ovoj tezi ne uspijevaju utvrditi istu i da se, zapravo, ne dovodimo pod bilo kakvu pozitivnu koncepciju u misli iz prvog lica, već da razmišljamo o sebi direktno i bez posredovanja koncepcije. To objašnjava zašto se čini otvorenim pitanjem, u slučajevima fisije, da je izvorna osoba identična s jednim od svojih nasljednika, dok je ono što je moguće jest da je ograničena pravom prirodom jastva na što se referira u misli iz prvog lica. Također uzgredno argumentiram da unutar treće interpretacije, prva premisa argumenta je nekonzistentna s nužnošću identiteta.

Ključne riječi: subjekt-tijelo dualizam, slučajevi fisije, misao iz prvog lica, transtemporalni identitet, Martine Nida-Rümelin

JUSTIFYING ONESELF

OPRAVDANJE SEBE

MARK PIPER
James Madison University

ABSTRACT

At present, the activity of justifying oneself is mostly discussed in psychology, where it is typically viewed as a negative or at least regrettable activity involving changing one's attitudes, beliefs, and feelings in order to minimize psychological threats arising from cognitive dissonance. Yet there is conceptual space, even a need, for an analysis of justifying oneself that is more content-neutral in nature. In this paper I provide such an analysis. Along the way I also briefly canvass some of the empirical work on self-justification in psychology and gesture towards issues surrounding the normative significance of the practice of justifying oneself.

Keywords: justification, self-justification, self-defense, commitment, internal justification, external justification, dialectical interaction

SAŽETAK

O aktivnosti opravdavanja sebe trenutno se najviše raspravlja u psihologiji gdje se na to tipično gleda kao na nešto negativno ili, u najmanju ruku, kao na aktivnost vrijednu žaljenja koja uključuje promjenu nečijih stavova, vjerovanja i osjećaja ne bi li se smanjile psihološke prijetnje što proizlaze iz kognitivne disonance. Unatoč tome, postoji konceptualni prostor, čak i potreba, za analizom opravdavanja sebe koja će po prirodi biti više sadržajno neutralna. U ovom radu nudim takvu analizu. Usput, također, ukratko istražujem neke od empirijskih radova o samoopravdanju u psihologiji i geste prema pitanjima koja okružuju normativno značenje prakse opravdavanja sebe.

Ključne riječi: opravdanje, samoopravdanje, samoobrana, predanost, unutarnje opravdanje, vanjsko opravdanje, dijalektička interakcija

A DEFENCE OF A RATIONALIST CONCEPTION OF PRACTICAL REASON

OBRANA RACIONALISTIČKE KONCEPCIJE PRAKTIČKOG UMA

GAL YEHEZKEL
The Sapir Academic College & The Open University of Israel

ABSTRACT

In this paper I attempt to refute the instrumental conception of practical reason, and thus defend a rationalist conception of practical reason. I argue that, far from merely playing an instrumental role, reason can be used by an agent to evaluate, that is, to approve or reject, final ends, which might be suggested by desires, and further to determine final ends independently of any desires, whether actual or potential, that the agent might have. My argument relies on an analysis of the concept of intention, and, more specifically, on the distinction between want and intention. I argue that the notion of an intentional action entails that reason can be used to evaluate and determine final ends.

Keywords: end, instrumental reason, intention, practical reason, reason

SAŽETAK

U ovom radu pokušavam osporiti instrumentalističku koncepciju praktičkog uma i, stoga, obraniti racionalističku koncepciju istoga. Argumentiram da, daleko od toga da igra samo instrumentalnu ulogu, agent može koristiti um za procjenjivanje, odnosno, ne bi li odobrio ili odbacio, krajnje ciljeve koji mogu biti predloženi od strane želja, i nadalje, ne bi li odlučio koji su krajnji ciljevi nezavisno od bilo kakvih želja, aktualnih ili potencijalnih, što ih agent može posjedovati. Moj se argument oslanja na analizu koncepta namjere i, točnije, na razliku između želje i namjere. Argumentiram da ideja intencionalne akcije podrazumijeva da um može biti korišten ne bi li se procijenilo i odlučilo o krajnjim ciljevima.

Ključne riječi: cilj, instrumentalni um, namjera, praktički um, (raz)um

"YES, THE THEORY IS ABSTEMIOUS, BUT . . .": A CRITIQUE OF YEHEZKEL

„DA, TEORIJA JE UMJERENA, ALI . . .": KRITIKA YEHEZKELA

REGAN LANCE REITSMA
King's College

ABSTRACT

This article is a critique of Gal Yehezkel's attempt to refute subjectivism about normative practical reasons, a school of thought inspired by Hume. Yehezkel believes reason, far from being, as Hume puts it, "the slave of the passions," has the normative authority to be a critic of basic desires and argues that subjectivism lacks the theoretical resources both to acknowledge this alleged truth and to analyze the distinction between wanting an outcome and intending to pursue it. I contend his refutation fails, largely because it operates with a strikingly attenuated view of the subjectivist theory.

Keywords: Hume, practical reason, normativity, rationalism, subjectivism, realism, constructivism, instrumentalism, basic desires, personal ideals, maieutic ends

SAŽETAK

Ovaj je članak kritika pokušaja Gala Yehezkela da ospori subjektivizam o normativnim praktičkim umovima, školu misli inspiriranu Humeom. Yehezkel vjeruje da je um daleko od, kako Hume kaže, *roba strasti*, da ima normativni autoritet koji mu omogućuje kritičnost naspram temeljnih želja, i argumentira da subjektivizmu nedostaju teoretski izvori ne bi li priznali tu navodnu istinu kao i analizirali razliku između željenja ishoda i namjeravanja da ga se postigne. Tvrdim da ovo osporavanje ne drži vodu ponajviše stoga što operira s nevjerojatno razblaženim pogledom na subjektivističku teoriju.

Ključne riječi: Hume, praktički um, normativnost, racionalizam, subjektivizam, realizam, konstruktivizam, instrumentalizam, temeljne želje, osobni ideali, majeutički ciljevi

Abstracts translated by:
Jelena Kopajtić, University of Rijeka, jelena.kopa@hotmail.com

INSTRUCTIONS FOR AUTHORS

Submission

Submitted manuscripts ought to be:

- unpublished, either completely or in their essential content, in English or other languages, and not under consideration for publication elsewhere. If it emerges that an accepted submission does not satisfy the previous conditions, the Editors reserve the right to reject it. If, due to the authors' failure to inform the Editors, already published materials will appear in EuJAP, the Editors will report the authors' unethical behaviour in the next issue. In any case, the Editors and the publisher will not be held legally responsible should there be any claims for compensation following from copyright infringements by the authors
- sent exclusively electronically to the Editors (eujap@ffri.hr) (or to the Guest editors in the case of a special issue) in a Word compatible format
- prepared for blind refereeing: authors` names and their institutional affiliations should not appear on the manuscript. Moreover, "identifiers" in MS Word Properties should be removed
- accompanied by a separate file containing the title of the manuscript, a short abstract (not exceeding 250 words), keywords, academic affiliation and full address for correspondence including e-mail address
- in American or British English
- no longer than 8000 words, including footnotes and references

Style

Accepted manuscripts should:

- follow the guidelines of the most recent Chicago Manual of Style
- contain footnotes and no endnotes
- contain references in accordance with the author-date Chicago style, here illustrated for the main common types of publications (T = in text citation, R = reference list entry)

Book
T: (Nozick 1981, 203)
R: Nozick, R. 1981. *Philosophical Explanations.* Cambridge: Harvard University Press.

Chapter or other part of a book
T: (Fumerton 2006, 77-9)
R: Fumerton, R. 2006. The Epistemic Role of Testimony: Internalist and Externalist Perspectives. In *The Epistemology of Testimony*, ed. J. Lackey and E. Sosa, 77-92. Oxford: Oxford University Press.

Edited collections
T: (Lackey and Sosa 2006)
R: Lackey, J. and E. Sosa, eds. 2006. *The Epistemology of Testimony*. Oxford: Oxford University Press.

Article in a print journal
T: (Broome 1999, 414-9)
R: Broome, J. 1999. Normative requirements. *Ratio* 12: 398-419.

Electronic books or journals
T: (Skorupski 2010)
R: Skorupski, J. 2010. Sentimentalism: Its Scope and Limits. *Ethical Theory and Moral Practice* 13: 125-136.
http://www.jstor.org/stable/40602550

Website content
T: (Brandon 2008)
R: Brandon, R. 2008. Natural Selection. *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. Accessed September 26, 2013.
http://plato.stanford.edu/archives/fall2010/entries/natural-selection

Forthcoming
For all types of publications followed should be the above guideline style with exception of placing 'forthcoming' instead of date of publication. For example, in case of a book:
T: (Recanati forthcoming)
R: Recanati, F. forthcoming. *Mental Files*. Oxford: Oxford University Press.

Unpublished material
T: (Gödel 1951)
R: Gödel, K. 1951. *Some basic theorems on the foundations of mathematics and their philosophical implications*. Unpublished manuscript, last modified August 3, 1951.

Final proofreading

Authors are responsible for correcting proofs.

Permission to republish

For permission to republish material in print or electronic format please contact the Editors (eujap@ffri.hr).

Subscriptions

A subscription comprises two issues. All prices include postage.

Annual subscription:
International:
individuals € 10
institutions € 15
Croatia:
individuals 40 kn
institutions 70 kn

Bank: Zagrebačka banka d.d. Zagreb
SWIFT: ZABAHR 2X
IBAN: HR9123600001101536455
Only for subscriptions from Croatia,
please add: “poziv na broj”: 0015-03368491

European Journal of Analytic Philosophy is published twice per year.

The articles published in the European Journal of Analytic Philosophy are indexed in The Philosopher’s Index and included in ERIH PLUS.