"WHY SHOULD I BE MORAL?" A CRITICAL ASSESSMENT OF THREE CONTEMPORARY ATTEMPTS TO GIVE AN EXTRA-MORAL JUSTIFICATION OF MORAL CONDUCT

by

Johnnie R. R. Pedersen

Dissertation submitted for the degree of M. Phil.

University of St. Andrews

September, 2006.

ACKNOWLEDGEMENTS

This dissertation could not have been completed, let alone begun, without the support (not least financially) of my parents – Frede and Else Marie. I am truly grateful for their support. Thanks also to my grandmother Jytte who, by letting me work at her house during my time in Denmark, in the winter of 2005/6, enabled me to work undisturbed while I was there.

Thanks to my supervisors Tim Mulgan and John Skorupski for invaluable discussion, comments and criticisms.

I am indebted to my friend Robert Pulvertaft for helping me out at various points with different aspects of my dissertational work. I wish him all the best, and hope that he will finish his own dissertation some day. I would also like to thank the rest of my friends for helping out, either directly or indirectly with the completion of this dissertation. I am particularly grateful for Emma Wilson's willingness to proofread the text.

Finally, I would like also to thank *Nordea Danmark Fonden* for financial support.

ABSTRACT

In this dissertation I consider three distinct attempts to answer the normative question "Why should I be moral?", all of which assume that a successful answer must be capable of arguing someone who is currently not motivated by moral considerations at all into becoming moral. I outline an argument against the possibility of doing so which relies on the distinction between agent-relativity and agent-neutrality, and which states that since morality essentially involves agent-neutrality and since failure to recognize the reason-giving force of agent-neutral considerations is not necessarily irrational, one cannot be argued into being moral. I then show how the approaches of Christine Korsgaard, as encountered in her *The Sources of Normativity*, Joseph Raz, as he puts it forth in "The Amoralist", and lastly, David Brink as he puts it forth in "Self-Love and Altruism", each in their different ways, fail in their attempts to argue someone into becoming moral.

TABLE OF CONTENTS

INTRODUCTION

Suppose you ask in some particular situation 'Why should I be moral? Why should I not perform the immoral act?', and I reply 'Because it would be wrong' would I – assuming that what I say is what there is most reason to believe – have answered your question successfully, or is something more required for that?

In this dissertation I offer a critical assessment of three distinct attempts to answer this so-called normative question, all of which assume that answering it requires not only showing that it would be wrong to perform the immoral act, but also that there is reason for any particular agent to make this consideration weigh with him. In this introductory section, I would like to say something general about the question itself, and present an argument to the effect that the approaches which I will be investigating – summarized under the heading 'Arguing someone into morality' – are bound to be abortive.

The normative question asks for a justification of moral conduct. When we ask for a justification of something, we ask for normative reasons, that is, considerations that count in favour of it.[1] It seems that three categories of reasons can be distinguished, and that being moral involves having reasons of each kind. Firstly, there are reasons for belief, for example believing the proposition that 'genocide is morally wrong'; secondly, there are reasons for acting, such as reasons for keeping a promise; and lastly there are reasons for feeling certain things. In each of these cases we may legitimately ask what considerations count in favour of believing, doing, or feeling the thing in question. This dissertation will focus on reasons to act, although *being moral* certainly involves certain dispositions for belief and feeling as well.

[1] Here I follow Scanlon (1998: 17): "I will take the idea of a reason as primitive. Any attempt to explain what it is to be a reason for something seems to me to lead back to the same idea: a consideration that counts in favor of it."

Next, it is important to distinguish between a *motivating* reason, i.e. a consideration some agent takes to count in favour of doing something, whether or not it actually does, and a *normative* reason, i.e. one that supports doing it, irrespective of whether the agent realizes it or not. Since the normative question is not asking for a descriptive account of what an agent actually believes, does or feels, but instead is concerned with what he ought to believe, do, or feel, the question should be understood in terms of normative reasons.[2]

Now, to the man in the street the answer to the normative question takes its most simple – and I would say, ultimately correct – form. If asked, for example, why the shopkeeper should not give his customers too little change even if he could do so with impunity, he will most likely say something like: Because it would be cheating them, because it would be taking unfair advantage, because it would be dishonest, or something akin to that. The fact that it would be cheating is the reason why he ought not do it – since that counts against doing the action – and to the vast majority this will be a quite satisfactory answer to the normative question.

To some, however, this kind of answer is not acceptable. Perhaps because they refuse to accept that normative reasons are metaphysically primitive entities that cannot be analyzed in terms of other concepts that do not include them. Such people seem to want to find a deeper element in normative thought which can tell us what the (ordinary) man in the street *really* means when he holds that an act ought to be avoided because (say) it would be cheating. They are looking for a reason for the reason, as it were: something by virtue of which the fact that an act would amount to cheating gives one a reason not to do it, such that someone who does not take that consideration to count against doing it, could come to see that it is a reason.

[2] However, it certainly seems correct that on a naturalistic construal, as J. S. Mill argued, we can only base our proof that something is desirable (that there is something counting in favour of it) on the observation that people in fact desire it (see *Utilitarianism*, ch. IV, parag. 3).

When the normative question is understood in this manner – as requiring that sort of reply – the most natural understanding of it viz.: 'Is there something counting in favour of acting morally?', has been rejected. Now the quest seems to be to find something by which this or that particular agent, irrespective of prior motivations, can be persuaded into believing what the man in the street already believes without further ado. In other words, we are no longer embarked on a metaphysical enquiry, but rather an epistemic one. The metaphysical question has already been answered satisfactorily – now we want to know how an agent can come to see this; whether he can come to realize that he ought to make this consideration weigh with him.[3]

How might one go about arguing someone into accepting what the man on the street accepts, without further reasons: that the fact that an act is morally wrong gives one a reason not to do it? Normally, one starts with the assumption that there are certain things which the agent is already committed to, which then, if the agent were to deliberate rationally on the basis of true beliefs alone, would make him realize that he is committed to acting on moral considerations. Hence, because of certain reasons, values, or commitments which he has already endorsed (or perhaps ought to endorse because of those endorsements), one brings to the fore certain sound argumentative steps which would get him (if rational) to accept the validity of morality. The three approaches which I shall consider in this dissertation are all examples of this.

On the face of it, it would seem that such attempts to argue someone into accepting moral demands must hinge on what specific reasons, values or commitments are in question. Surely, whether or not we can successfully attain moral reasons or values

[3] As Foot (1972) pointed out, it seems entirely consistent to recognize the existence of categorical moral norms, but still ask why one should comply with those norms. The kind of approach to the normative question that I am outlining here seems to want to address exactly this sort of moral externalist challenge, by arguing that there is reason for such an agent to recognize the reason-giving force of those norms.

from certain other reasons or values must depend at least in part on what those other reasons and values are.

The three approaches which shall be considered here are all seem to assume that non-moral commitments are enough to make the argument succeed. They assume this not without good reason, though. For since it is presupposed (because of the way in which the normative question is understood) that the agent does not already believe that moral considerations have reason-giving force, it would be circular to appeal to the reason-giving force of those considerations when attempting to justify them. The project is not simply that of pointing out to some agent that acting in this manner would indeed be unfair, base, or cruel, since appeal to such properties would not be enough to motivate someone who is not already motivated by moral considerations.

This is what I understand by the task of 'arguing someone into morality': showing, on the basis of some non-moral considerations, that there is reason to be moral. Because of this very feature of the approaches, however, I do not believe that they can be successful, since I take it to be impossible to establish the reason-giving force of morality on the basis of something extra-moral.

To clarify this, we might start by distinguishing between two different conceptions of how normative reasons have to be related to the agent. Bernard Williams has argued that there are only what he calls internal reasons, that is, reasons that are internal to what he calls an agent's "subjective motivational set" (1981: 102) – abbreviated 'S'. According to Williams, all "reasons-sentences", that is, utterances of the form "'A has a reason to Φ' or 'There is reason for A to Φ' (where 'Φ' stands for some verb of action)" can be interpreted in two different ways:

> "On the first, the truth of the sentences implies, very roughly, that A has some motive which will be served or furthered by his Φ-ing, and if this turns out not to be so the sentence is false: there is a condition relating to the agent's aims, and if this is not satisfied it is not true to say, on this

interpretation, that he has a reason to Φ. On the second interpretation, there is no such condition, and the reason-sentence will not be falsified by the absence of an appropriate motive." (1981: 101)

The first interpretation of reasons-sentences expresses what he calls internal reasons, whereas the second expresses what he calls external reasons. Williams' claim is that the only tenable interpretation of reason-sentences is the internal one. In other words, in order for something to be a reason for A to Φ, it must be true that A has some motive which will be served or furthered by him Φ-ing.[4] Let us call this view (W):

[4] Williams presents two arguments in support of this. The first challenges the externalist about reasons to account for the truth-conditions of statements like 'There is reason for A to Φ'. According to the externalist it may be true that 'There is reason for A to Φ' despite the fact that Φ-ing could serve no motive in A's S. But "[w]hat is it", Williams asks, "the agent comes to believe when he comes to believe he has a reason to Φ? If he becomes persuaded of this supposedly external truth, so that the reason does then enter his S, what is it that he has come to believe? This question presents a challenge to the external theorist" (1995: 39). What, in other words, is the meaning of the claim 'There is reason for A to Φ' on the externalist construal, when the agent comes to accept that he has reason to Φ? And of course the externalist cannot explain its meaning by saying that the agent comes to believe that Φ-ing would serve a motive of his. If he cannot say that the agent comes to believe that Φ-ing would serve a motive in his S – and this is the second challenge – then how can the reason serve as an explanation of what he does if and when he Φs? According to Williams, externalism clashes with "the interrelation of explanatory and normative reasons" which states that "[i]f it is true that A has a reason to Φ, then it must be possible that he should Φ for that reason" (1995: 39). But if A cannot or does not see reason r as a reason for Φ-ing himself, then presumably he will not Φ for that reason. But by the principle of the interrelation between explanatory and normative reasons, if it is impossible that A should Φ for reason r, then it cannot be true that A has *that* reason to Φ (which in turn is what externalism claims).

(W) A has a reason to Φ if and only if A has some motive which will be served or furthered by his Φ-ing.[5]

Now it seems that accepting (W) would clearly render it impossible to argue someone into accepting moral demands (in the sense under consideration here). This becomes evident if we consider the case of the rational egoist. The rational egoist is someone who recognizes only agent-relative reasons stemming from self-interest: he thinks that the only considerations which are reasons to act are ones which show that acting in this or that way would consist, or result, in a benefit to him – either directly, or indirectly by benefiting his intimates, advancing his causes, or achieving his aims. As I shall go on to argue, however, a commitment to morality by its very nature involves a commitment to the existence of not only agent-relative, but also to agent-neutral reasons to act (concepts which will be defined formally shortly). For example, the fact that some act is, say, the killing of an innocent individual, x, is, when considered from the moral point of view, a reason not to perform the act, regardless of who x is – me or anyone else. But since there is nothing in the logic of (W) which requires that an agent recognizes agent-neutral reasons for action, an agent cannot be argued into accepting morality. And this in turn shows that any attempt to argue someone into accepting moral demands on the basis of (W) is foredoomed to failure.

To make these points clearer, we must first define the distinction between agent-relative and agent-neutral reasons:[6]

[5] Williams takes account of cases in which an agent has motives based on false beliefs by excluding them from S, such that, if an agent wants a gin and tonic, for example, he does not have reason to drink the stuff in this glass – which is petrol – although he believes it is gin (1981: 102; 1995: 36). Importantly, he also stresses that deliberation can change the agent's S: "[w]e should not […] think of S as statically given. The process of deliberation can have all sorts of effect on S, and this is a fact which a theory of internal reasons should be very happy to accommodate" (1981: 105).

Agent-relative reason$_{df}$: R is an agent-relative reason for an individual agent x to do A if and only if R (the consideration that counts in favour of doing A) essentially involves a token-reflexive backward reference to x.

Agent-neutral reason$_{df}$: R is an agent-neutral reason for an individual agent x to do A if and only if R (the consideration that counts in favour of doing A) involves no essential token-reflexive backward reference to x.[7]

This distinction is meant to capture the common-sense idea that there are some things which there is reason for people to do, for the very reason that they are who they are, as it were. On the other hand, there seem to be other actions which there is reason for people to perform, independent of who they are. So, for instance, the fact that a particular human being is your spouse seems to be a reason for you to prefer to save her, rather than a complete stranger, if both of them are in danger of drowning. That is simply because she is your spouse. This is the sort of intuition that the definition of agent-relative reason is intended to incorporate: the consideration that counts in favour of saving the individual who is your spouse, is the fact that that individual is your spouse.

---

[6] Thomas Nagel was the first to introduce this distinction (Nagel 1970). However, he did not use the terms 'agent-relative' and 'agent-neutral', but instead 'subjective' and 'objective' (see his definition of this distinction on p. 91, which hinges on an understanding of the universality of reasons stated on p. 47). Derek Parfit (1984) introduced the 'agent-relative'/'agent-neutral' distinction, and he took it to be equivalent to Nagel's 'subjective'/'objective' distinction (see p. 143). Nagel (1986: 152-3) then adopted this terminology and rejected his earlier one.

[7] Or simply: R is an agent-neutral reason if and only if it is not an agent-relative one.

But it seems equally true that there are certain acts that there is reason for everyone to perform, no matter how they are related to the action or its outcome. For example, many would agree that there is reason for anyone to save somebody who is drowning. In this case there is no essential reference to any particular individual, and hence the fact that someone is drowning is an agent-neutral reason.

Related to this, is the distinction between agent-relative and agent-neutral values, which I propose we define in terms of the previous distinction:

Agent-relative value $_{df}$: V is an agent-relative value (a value for a particular agent, x), if and only if there is a agent-relative reason for x to Φ.

Agent-neutral value $_{df}$: V is an agent-neutral value, if and only if there is a agent-neutral reason to Φ.

Where Φ-ing refers to whatever response is appropriate to value V. For example, valuing good health gives me (according to the results of scientific research) reason to 'eat 5 a day', exercise, avoid stress, not consume too much alcohol, and a whole range of other things. In other words, if good health is of value to me, i.e. an agent-relative value, that means that there is reason for me to do those things, because that would be (according to scientific research) the rational way to react, since, supposedly, taking these precautions is a way of achieving my goal. Likewise, claiming that good health is an agent-neutral value, means claiming that there is reason for everyone to do the things in question.

Bearing this distinction in mind, the argument against the possibility of answering the normative question in the positive can be stated in this way: First we assume that Williams' internalism is true, and then we argue along these lines:

(1) There is a sound difference between agent-relative and agent-neutral reasons.

(2) Showing that there is reason to be moral requires showing that there is reason to be motivated by moral considerations (e.g. that it is one's duty to perform this act, or that x needs it or is benefited by me doing it).

(3) Some moral considerations are essentially agent-neutral.

Therefore ((2), (3)): (4) Showing that there is reason to be moral requires showing that there is reason to be motivated by certain agent-neutral considerations.

(5) Nothing in the logic of (W) entails that rational agency requires being motivated by agent-neutral considerations.

Therefore ((4), (5)): (6) Accepting (W), does not necessarily commit one to being moral.

There is nothing inconsistent in refusing to recognize agent-neutral reasons. Therefore, one cannot be argued into accepting morality if (W) is correct.

It might be objected that it would be inconsistent for someone to fail to recognize moral reasons, since, surely, *any* agent would want others to conduct themselves according to the demands of morality in their interactions with oneself, since one would obviously benefit from their doing so. If you believe that others have reason, say, not to deceive you, then clearly consistency requires you to believe that you have reason not to deceive others.

It is generally agreed, the objection goes, that reasons are universal in the sense that if the fact that p is a reason for some particular agent to Φ in circumstance C then for any x, p is a reason for x to Φ in circumstance C. Thus, if you believe that the fact that p is an act of deceit against you constitutes a reason for everyone not to perform act p,

then the universality of reasons commits you to believing that the fact that act q is an act of deceit against anyone else is a reason for you not to do it, assuming that the circumstances are similar. Generalizing from this we deduce that if you believe that the fact that an act is to your disadvantage is a reason for everyone else not to do it, you ought to believe that the fact that an act is to anyone's disadvantage is a reason not to do it – one cannot consistently believe that the reasons which apply to oneself do not apply to others, in similar circumstances.

This charge of irrationality is correct as far as it goes. However, when we take the aforementioned double meaning of 'reason' into account, it becomes clear that one can consistently hold that the fact that some act constitutes or results in a benefit to oneself constitutes a reason to perform it.

When (W) is interpreted solely in terms of agent-relative reason, there is nothing irrational about refusing to recognize agent-neutral reasons. To see this, let me introduce the rational egoist, i.e. someone who conducts himself according to, and solely according to, the Principle of Rational Egoism (hereafter "PRE"):

PRE: (for any individual agent x and for all acts y) (if and only if doing y constitutes or results in a benefit to x, then x has a reason to do y).

In other words, the rational egoist is someone who takes the fact that some action constitutes or brings about a benefit for him to be an agent-relative reason for acting, and so not necessarily to be a reason for anyone else besides him. A person endorsing PRE would have to admit, then, that other people do not necessarily have any reason to benefit him – that all depends on whether his well-being as a part of their own, which it might be if they coincidentally were his friends, or parents, for example. Since, as the above objection correctly supposes, rationality imposes a requirement of

universalization, it would be (trivially) irrational for one to hold purely – as does the Irrational Egoist – the Principle of Irrational Egoism (henceforth "PIE"):

PIE: (for any individual agent x and for all acts y) (if and only if doing y constitutes or results in a benefit to me, then x has a reason to do y).

PIE cannot be universalized because it contains the token-reflexive, or rigid designator 'me'. Surely there is nothing special about 'me' which could account for the rationality of benefiting only 'me': a lot of other people, if not the majority of them, are similar in relevant respects.

Now, the above objection, that it is irrational for someone to accept the existence of agent-relative reasons, while not believing in the existence of agent-neutral reasons, seems to be based on the false assumption that PIE is the only principle that allows one to deny the existence of agent-neutral reasons to benefit other people. But clearly both PRE and PIE permit one to deny this, and since PRE is not irrational, the objection fails; one can in fact fail to accept the existence of agent-neutral reasons, and believe that everyone's well-being matters, because one can refuse to believe that it matters rationally to anyone besides the person whose well-being is in question. Therefore, there is nothing in the logic of agent-relative reasons that (in and of itself) commits one to the validity of agent-neutral reasons.

An agent subscribing to PIE (as his only principle) claims that there is reason for everyone to benefit him. However the requirement of universalization imposed by rationality has the consequence that holding such a position would be irrational, unless – as there is good reason to suspect cannot be done – the agent endorsing PIE (as his sole principle), can come up with good reasons why others' well-being are relevantly different from his own in significance.

On the contrary, PRE *is* a principle of rational agency. The rational egoist endorsing PRE claims merely that *his* well-being gives *him* a reason to act, that is, that his well-being gives him an agent-relative reason to act – not that others have an agent-neutral reason to do anything about his well-being. Since he does not deny that others likewise might have agent-relative reason to take care of *them*selves, his position is entirely consistent. The upshot of this discussion then, is that there is no hope of showing – merely on the basis of (W) – that immoral conduct is irrational. Hence, since immoral conduct is consistent, neither can one cannot be argued into accepting morality.[8]

If the preceding argument is correct then we cannot start with just any motivational set, in particular one that does not include agent-neutral motivations, and argue someone into morality. Since morality essentially involves a commitment to agent-neutral reasons, and since one may consistently fail to recognize the motivational force of such reasons, there is no way this project can succeed.[9]

---

[8] An illustration of the impossibility of arguing someone into accepting moral demands is Henry Sidgwick's "Dualism of the Practical Reason" (Sidgwick 1907: xxi), which states that according to ordinary intuitions there are two irreducible, ultimately right-making principles: what he calls the principles of "Egoistic and Universalistic Hedonism" (ibid, 11). The principle of egoistic hedonism states that rational action is action done for the sake of one's own pleasure, whereas universalistic hedonism claims that actions performed out of concern for everyone's pleasure are rational. Since both of these considerations have the ability to render an act rational, and granted the fact that moral and prudential considerations can potentially conflict (recognized by Sidgwick, ibid. 9-10), it is impossible to show that immoral conduct is necessarily irrational. For useful discussions of Sidgwick's dualism of practical reason, see J. Mackie (1976) and J. B. Schneewind (1977, esp. ch. 13).

[9] As C. Korsgaard argues in her (1986) it is consistent with Williams' internalism that there may be categorical reasons stemming from pure practical reason, such that categorical moral reasons can be derived from every rational deliberator's S. Whether or not it may be possible to derive the demands of morality from pure practical reason, I shall set this issue aside, since none of the three subsequent attempts embarks on the project of doing so. Although Korsgaard, as we shall see in chapter 1, does rely on Kantian-like arguments in her attempt to show that there is reason to be moral, she relies on

Now I am not claiming that (W) expresses the correct conception of practical rationality. However it is still holds true that one cannot be argued into accepting moral demands even if one denies the truth of (W). And for a parallel reason: just as we cannot make any reference to something in the agent's S when we are stating why he should be moral, since by assumption there is no moral motivation there in the first place to appeal to, so we cannot – because of the very way the normative question is understood – appeal to any moral fact when stating the external reason.

Accepting internalism, we are debarred from making reference to any moral motive in the agent's S, because we would then not be arguing him into accepting morality in the first place, but merely making him aware of motivation he already had. On externalist grounds we cannot appeal to the reason-giving force of moral considerations either, since that is exactly what is in contention. But if we cannot appeal to moral considerations in constructing our argument, neither can we make any appeal to that part of morality which consists of agent-neutral reasons for action. Since this is so, and since, further – as the case of the rational egoist illustrates – there is no analytical connection between agent-relative reasons, values, commitments, etc., and agent-neutral ones, one cannot be argued into morality on the basis of externalism about reasons either.

---

the distinctly un-Kantian idea that our reasons are something we choose, instead of constraints on our choices, which we acknowledge by our exercise of practical reason. Whether moral obligations can in fact be derived from pure practical reason is a question which, although extremely interesting, falls beyond the scope of this dissertation. I hope to be able to discuss this at another occasion. Let me just note here that Kantians normally also stress the importance of autonomy, i.e. of the agent being able to tell for himself what he should do, without having to listen to any external authorities. So, by implication: insofar as the rational egoist – despite the fact that pure practical reason demands it – is unable to tell that there is reason for him to do the morally right thing, it is not altogether clear whether they would claim that there is a normative reason *for him* to do it.

That is what seems to me to be the truth, and in this dissertation I want to demonstrate in detail how three different attempts to show how someone can in fact be argued into accepting morality, all fail.

The first two of these – that of Christine Korsgaard and that of Joseph Raz – are distinctively modern in outlook in that they take impartiality to be a necessary feature of morality. In other words, they grant that morality sometimes requires that one's actions and attitudes not be biased by considerations concerning who will be benefited or harmed by one's actions or attitudes. Hence, according to these two attempts, being committed to morality goes hand in hand with a commitment to agent-neutral reasons to act.

As a result, in their attempts to answer the normative question they argue that there is agent-neutral reason for us to be concerned about other people. However, they openly admit that we might recognize only agent-relative ones. Since, however, they openly admit that someone who does not believe in the reason-giving force of agent-neutral considerations can be argued into accepting moral demands, these authors face the seemingly unbridgeable abyss between agent-relativity and agent-neutrality, which I have argued for above.

The third attempt – a sophisticated version of eudaimonism, advanced by David Brink – does not share the view that morality is essentially agent-neutral in part. Instead, it holds that being moral is in the agent's own self-interest, and attempts to show how we can give an egoistic justification of the whole of morality solely on this basis.

Hence, corresponding to the two diverging views on what morality is about, we get two different kinds of approaches to the normative question: The first two attempt, in different ways, to bridge the gap between agent-relativity, and agent-neutrality, and the third tries to show that, as a matter of fact, an agent-relative consideration,

namely concern for one's own well-being, will make it rational for one to comply with moral demands.

The two attempts to bridge the gap between agent-relativity and agent-neutrality will be assessed in the first two chapters: That of Korsgaard in the first and that of Raz in the second. Both of these philosophers assume that morality involves a commitment to impartiality, such that one necessary feature of morality is that some moral reasons apply to agents not because of agent-relative considerations, but because of agent-neutral ones.

However, whereas Korsgaard tries to show that possession of practical reason – that is, the capacity to recognize, weigh, and resolve on the basis of reasons – goes hand in hand with moral obligations, Raz argues that certain empirical facts about a normal human life – specifically that they typically involve deep interpersonal relationships such as friendships – commits us to the value of those people, the recognition of which in turn, more or less, would commit an egoist to the value of all people. I shall argue that both of these attempts fail because of a failure to appreciate fully the agent-relative/agent-neutral distinction.

Historically there seem to be two major ways to go about arguing that we have an egoistic, agent-relative reason to be moral. The first strategy is adopted by eudaimonism, which attempts to argue that human flourishing, or less botanically, human well-being, is intimately tied up with living according to moral demands, since the well-being of others is an intrinsic component of one's own well-being. By contrast, the second type of strategy, Hobbesian contractualism, denies that the well-being of others is an intrinsic part of the agent's own well-being, but rather regards morality as a set of mutually beneficial rules of conduct. Thus, although self-interest does at times give the agent a *pro tanto* reason to act contrary to what is morally required, it just so happens that *as a matter of fact* there is always *conclusive* reason to do what is morally demanded. Both of these strategies however, share the intuition

that morality is essentially agent-relative in the sense that it is founded on agent-relative or egoistic reasons, i.e. reasons to benefit oneself.

Eudaimonism holds that an agent's practical reasoning should be guided by an objectively correct conception of eudaimonia – that is, the ultimate end any given rational agent has reason to pursue is one's own happiness. Furthermore, since eudaimonists regard possession and exercise of the moral virtues as an essential part of eudaimonia, they are committed to the view that the possession and exercise of moral virtues is an essential part of happiness. Hence, in their attempts to establish that one has reason to be moral, the eudaimonists proceed in the same way as the contractualists: they attempt to show that adherence to the demands of morality – possession and exercise of moral virtue – will contribute to making one's own life better. But whereas eudaimonists think that the morally good life is an essential part of the good life simpliciter, Hobbesian contractualists tend to think of each person's life as something distinct, and of moral duties as a set of mutually beneficial means to safeguarding the agent's own self-interest.

The manner of addressing the normative question, however, is basically the same: first a prudential conception of practical reasoning is assumed, and then what there is agent-relative reason to do is shown to be identical with what there is moral reason to do, by resorting to certain facts – in the case of eudaimonism, about the nature of human flourishing, and in the case of contractualism, about the social benefit of heeding certain rules of conduct.

Brink has proposed an approach to the normative question which draws on the eudaimonism of Plato and Aristotle, together with some insights from the philosophy of the British idealist T. H. Green. Like Raz, Brink also focuses on relationships with intimates. But he does not attempt to argue that the value one places on such relationships commit one to the agent-neutral value of mankind at large; rather, he argues that what makes relationships with intimates agent-relatively

good is also present – although to a lesser extent – in one's relationship to the rest of mankind. Hence one has an agent-relative reason to benefit them, in the same way as one has reason to benefit one's friends. I shall argue, however, that this approach does not withstand scrutiny, since it is unable to account for one essential, agent-neutral element of common-sense morality, namely justice.

Thus, my treatment of these three attempts will suggest a conclusion which will be a negative and – given the above discussion – unsurprising one, namely that one cannot by way of abstract philosophical argument be argued into accepting moral demands. This is because, ultimately, it is not necessarily irrational to refuse to recognize the motivating force of moral obligations. Since there is nothing internal to this sort of position, which could function as an argumentative basis on which we could reach the agent-neutral part of morality, the project of arguing someone into morality must be abortive.

1. KORSGAARD'S APPROACH: THE APPEAL TO INTEGRITY

In *The Sources of Normativity* Korsgaard puts forth an interesting approach to the normative question, which I will investigate in this chapter.[10] Before embarking on this task, it is necessary to shed some light on what the normative question, as understood by Korsgaard, is. In particular, I shall point out that Korsgaard's understanding of the normative question commits her, if her subsequent argument is to be at all cogent, to providing an answer as to why somebody who does not presently recognize moral demands as having any normative force on him has reason to become motivated by them. In other words, it is incumbent upon her to divulge to us considerations which would amount to a conclusive reason for someone external to moral practice and discourse to become a moral agent. By holding her to this commitment, which I argue that she is committed to in section 1.1, I will, after having stated and explained her main argument in section 1.2, show, in section 1.3 that it in fact falls short. As we shall see, Korsgaard tries to argue that the agent-relative/agent-neutral distinction is bogus in practice, because we can always 'intrude' ourselves into each other's minds, thus making the other feel the reason-giving force of moral considerations. I shall argue that Korsgaard's argument presupposes what it wants to establish. Besides this main critique, I shall argue that categorical moral reasons are incompatible with Korsgaard's voluntarism.

1.1 KORSGAARD ON THE NORMATIVE QUESTION

What does Korsgaard understand by the normative question? According to her, the question stems from the quest for a 'foundation' of morality – i.e. the search for something in virtue of which the existence and objective validity of moral obligations can be justified.

[10] Unless I indicate otherwise, all references in parentheses will be to this book.

When we ask for a philosophical foundation of morality we are not simply asking for an explanation as to why we are in fact moral, for instance by pointing out that we are by nature social creatures, or that our moral dispositions are results of an ongoing process of socialization, habituation and education. Rather, she explains, we want to understand the fact that ethical standards are *normative* to us, or what justifies the fact that these claims can exercise normative force on us.

"When we seek a philosophical foundation for morality we are not looking merely for an explanation of moral practices. We are asking what *justifies* the claims that morality makes on us. This is what I am calling 'the normative question'." (p. 9-10)

Merely explaining why we do, in fact, heed moral demands by citing various circumstances in our upbringing or social natures would not achieve this aim. For, once we have explained why we are actually following moral exigencies, one may still legitimately ask whether we are justified in doing so.

We may, for instance, explain A's conviction that it is wrong to break a promise in terms of his good upbringing, or in terms of his fear of eternal damnation, but that does not in and of itself justify his conviction. Merely unravelling why we are in fact moral would not justify moral demands themselves, and hence it does not constitute a satisfactory answer to the normative question.

So by posing the normative question, Korsgaard asks for a justification of the authority of moral demands – their claims on us, as she says. But one may well wonder what the extension of 'us' is here, for morality certainly seems to most of us to be making claims on *all* of us, whereas one can observe that only some of us are regularly and reliably motivated to act accordingly. Does Korsgaard intend her justification of moral demands to cut across both of these groups?

The rationale behind this question is that it appears to be one thing to establish, by way of philosophical argument, the normative force of moral judgements to people

who are, as it were, internal to the moral practice – who already feel the force of moral demands – but quite another to do so to people who are external in this sense. If one is a sentimentalist, say, and believes that all practical reasons are affectively grounded, then it would be relatively easy to show that one is justified in doing one's moral obligation if one is internal to morality, whereas it would be all the more difficult to show – vis-à-vis the argument put forth in the introduction – that someone who is not predisposed to follow moral demands, has a reason to do so.

In other words, the normative question might be asked from two completely different perspectives – from, as I shall say, either the internal or the external point of view. It may be asked either from the point of view of the moral person, who considers heeding moral demands a part of who he is (or a part of his S), or from the point of view of someone who, like the rational egoist, feels no such commitment. Accordingly, if we are to assess the soundness of Korsgaard's argument, we need some clarification as to whether 'us' is intended to cover only moral agents, humanity at large, rational agents.

Initially it is not quite clear whether Korsgaard intends her argument to supply someone internal or external to the moral point of view with an answer. She avows three criteria which an answer to the normative question must fulfil in order to be successful.[11] However, as I shall point out, although they seem to support the interpretation that Korsgaard only regards a justification as possible from the internal point of view, her own subsequent discussion and dismissal of other approaches, supports the opposite conclusion that she thinks a justification is possible from the external perspective.

The first of Korsgaard's criteria states that the answer must succeed in addressing the agent who asks it in the first person – that is, it must provide an answer to the question: 'Why must *I* do what morality requires?' Now, of course, a given answer

[11] See pp. 16-18.

may not succeed in the case of somebody who challenges moral demands altogether. For as Korsgaard makes clear,

> "He might be insincere and contentious; he might be looking for a way to evade his duty, rather than asking the question because he really wants to know. For this exercise to work, we have to eliminate these possibilities, and imagine that this […] agent is sincere and reasonable, and does really want to know." (p. 16)

This seems to indicate that insofar as the agent is sincere and reasonable, then Korsgaard would say that, a successful answer to the normative question must be capable of supplying him with a reason which would convince him of the normativity of moral demands. But Korsgaard's statement does not necessarily amount to the strong proposition that irrespective of whether an agent is internal or external to moral discourse, necessarily, if he is sincere and reasonable and really wants to know, a successful answer to the normative question will supply him with reason to adhere to moral demands.

On the other hand, such a commitment is not excluded by what Korsgaard says, for there seems to be no reason why someone external to moral demands might not be sincere and reasonable whilst still wanting to know whether he should become moral. It would be a very strong requirement indeed, because it would seem to require that we bridge the gap between moral conduct and rational conduct.

However, it is only if we equate 'reasonable' with 'rational' in the above quote, that the criteria make her committed to this strong claim. Substituting 'rational' for 'reasonable', in the above quote would make Korsgaard committed to the proposition that merely being rational, and sincere (which I take to be equivalent to 'willing to take into account practical reasons without prejudice'), would be enough for a successful answer to the normative question to succeed.

If, on the other hand, Korsgaard has something else in mind by 'reasonable' then she would fall short of a commitment to bridging the gap between moral and rational conduct. Normally, these terms are used interchangeably, but there is a Rawlsian tradition of distinguishing them, and understanding rational action as action instrumental to the agent's self-confined aims, whereas reasonable action is taken to involve the idea that the action has to be justifiable to others – something to which they could consent. If we interpret Korsgaard's use of 'reasonable' as carrying this latter meaning, then she would in effect only be intending her argument to apply to someone who is already interested in justifying his actions to others, that is, someone internal to morality. Because Korsgaard does not make any effort to distinguish between the two concepts, it is impossible, merely on this basis, to determine what she takes a successful answer to the normative question to be.

As I pointed out, Korsgaard states three conditions which she takes to be jointly necessary for an acceptable answer to the normative question, and so far only the first (that the agent asking it must be reasonable and genuinely seeking an answer) has been considered. The second is a requirement of transparency, i.e. that the knowledge of what justifies an agent in acting as he is morally required to must be capable of actually making him believe that his actions are justified. This condition seems irrelevant when it comes to determining what Korsgaard takes a successful answer to be, since it cuts across the distinction between being internal and external to morality.

Thirdly, however, it must, Korsgaard remarks, "appeal to our sense of who we are, to our sense of our identity." (p. 17) If this is so, then someone attempting to come up with an answer to the normative question would not be committed to providing someone external to moral demands with a satisfactory answer, if there is one to be had, for he would, trivially, not regard being moral as a part of his identity.

Therefore, neither would an attempt which failed to apply to him render the approach to the normative question unsatisfactory by Korsgaardian standards.

There are problems with this interpretation of the third condition, however. Korsgaard's argument for it seems to render it implausible that an agent will seriously ask the normative question in the first place. It goes as follows: we know from experience that morality often demands self-sacrifice, and sometimes even death. If, however, Korsgaard says, we are to be prepared to go to our deaths for the sake of morality, then a successful answer to the normative question

> "must show that sometimes doing the wrong thing is as bad or worse than death. And for most human beings on most occasions, the only thing that could be as bad or worse than death is something that for us amounts to death – not being ourselves any more." (p. 18)

But if the agent already considers it to be a loss of self, and hence something worse than death, would it make sense for him to ask the normative question, 'Why should I be moral?' in the first place? It seems not, for surely he would not want to go through something worse than death if he can get away with something less bad?

It would make sense though on the assumption that we are merely facing the challenge of making the person posing the normative question aware that being moral is indeed a part of her, such that she would not be herself anymore if she did something morally bad. We might do this by prompting her to exercise her imagination. If for example somebody believes that it is morally wrong to hurt other people, we might, by making her see that she hurts someone by breaking her promise to him, make her motivated not to break it.

So we must conclude that Korsgaard's formulations of the three criteria do support interpreting her as claiming that a successful answer to the normative question should only be expected to answer someone internal to moral demands, although she has not yet claimed that it should not answer someone external to morality.

However, as her subsequent discussion and dismissal of other approaches to the normative question makes clear, she does seem to demand that a successful answer to the normative question be capable of persuading someone external to morality that he has to join. For instance, she thinks voluntarism – according to which the normativity of moral obligations is nested in the command of someone who has authority over the moral agent – cannot appeal to morality itself in justifying this authority, since that would be circular.[12] Because she rejects this position on the ground that it makes reference to morality, it seems she must be supposing that a satisfactory answer must avoid making such an appeal.

The same line of critique is launched against realism, and early twentieth-century rational intuitionism.[13] Realism in particular answers the normative question by claiming that there are certain intrinsically normative entities, mostly actions, which in turn – because of their very nature – account for the normativity of moral demands. But Korsgaard thinks that

> "when the normative question is raised, these are the exact points that are in contention – whether there is really *anything* I must do, and if so whether it is *this*. So it is a little hard to see how realism can help." (p. 34)

This clearly suggests that she believes that a successful answer to the question must be satisfactory to someone external to morality.

Later she writes, also against the realists:

> "If someone finds that the bare fact that something is his duty does not move him to action, and asks what possible motive he has for doing it, it does not help to tell him that the fact that it is his duty just is the motive. That fact isn't motivating him just now, and therein lies his problem. In a

[12] Cf. p. 30.

[13] Cf. p. 32.

similar way, if someone falls into doubt about whether obligations really exist, it doesn't help to say 'ah, but indeed they do. They are *real* things'. Just now he doesn't see it, and therein lies his problem." (p. 38)

Again, this clearly indicates that a successful answer to the normative question must make someone who does not presently recognize the normative force of moral considerations be motivated to act on the basis of them.

To Korsgaard, answering the normative question is not just a matter, as H. A. Prichard (1912: 28) thought it often was – at least in those cases where the question makes sense – of pointing out to someone that there really is moral reason to perform this action, by giving a more comprehensive or precise statement of the action, e.g. that something is not the giving of a present to A, but repaying A by giving him a present, which in the right circumstances may change an agent's view on whether the action ought to be done or not. To Korsgaard on the contrary, it is a matter of supplying "someone who has fallen into doubt about whether moral requirements are really normative" (p. 38) with an answer that could bring one to be motivated by such facts, where one is not so motivated in the first place.

As I said *en passant* above, Prichard (1912) thought that the normative question, when understood the way Korsgaard poses it, is senseless. The argument he offers for this point is that such an understanding of the question leads to the following dilemma: if the answer is a moral reason, e.g. 'it is your duty', then what needs to be established will be needed, i.e. that there is reason to do one's duty. If, on the other hand, the reason is non-moral, e.g. that it promotes your self-interest, then the answer becomes irrelevant because the reason why one should do it is that it is morally right to do it. Showing that it is in your self-interest to perform the moral action may make you *want* to do it – it does not show that you ought to do it.

Korsgaard is certainly familiar with this dilemma, but, disappointingly, she does not discuss how it relates to her own approach to the normative question.[14] Although Prichard was an intuitionist, and although Korsgaard rejects intuitionism alongside realism, it still seems necessary to consider Prichard's dilemma; for there is no necessary connection between Prichard's intuitionism and his dilemma.

What Korsgaard says about realism (and supposedly intuitionism as well) is that

> "it refuses to answer the normative question. It is a way of saying that it cannot be done. Or rather, more commonly, it is a way of saying that it need not be done. For of course if I *do* feel confident that certain actions really are required of me, I might *therefore* be prepared to believe that those actions are intrinsically obligatory or objectively valuable, that rightness is just a property that they have." (p. 39)

But because Korsgaard assumes that we have to be able to argue someone external to morality into accepting moral demands, we cannot use those objective values to argue for the normativity of obligations, since the belief in the existence of these values itself presupposes the belief in the obligatory nature of certain moral actions, something which the person external to morality denies.[15]

However, it seems to me that this argument is clearly lacking in cogency, especially because she has not argued for the essential premise – denied by Prichard and others – that it is in principle possible to argue someone into accepting moral demands. Korsgaard seems just to assume that this is possible, and hence she seems to ignore the obstacle Prichard's dilemma constitutes. On the other hand, maybe she just thinks it is worthwhile to try to find an external justification for moral demands – to see where our thoughts on the matter will lead us. Whether Korsgaard succeeds in her attempt to show Prichard's contention – as well as the arguments which I

[14] Cf. p. 32.

[15] Cf. p. 40.

presented in the introduction – to be wrong will be revealed in the course of this chapter.

For now I conclude that Korsgaard's understanding of the normative question, 'Why should I be moral?', commits her to providing an answer to it which would satisfy somebody regardless of what his present motivations are. Although initially it proved difficult to pin down, on the basis of Korsgaard's three criteria of a successful answer to the normative question, whether she thought that an answer must satisfy someone internal or external to morality, her own critique of other approaches to the question reveal that she must be committed to persuading someone external to morality if her own attempt is to succeed. In the following I will hold Korsgaard to this quite strong commitment.

1.2 THE ARGUMENT

What Korsgaard professes to argue is that

> "if we take anything to have value, then we must acknowledge that we have moral obligations." (p. 92)

This is indeed an astonishing inference, for it involves, as I shall explain, a move from agent-relativity – that something is valuable to someone – to agent-neutrality, i.e. that there is reason for everyone to value it. In this section I would like to put forth and explain the steps making up the argument behind it. I thus postpone the main discussion and assessment of the argument to section 1.3, where I consider what I take to be the most crucial step: the attempt to bridge the gap between agent-relative and agent-neutral values. But for now I merely wish to summarize and explain the main steps in Korsgaard's argument.

Basically, Korsgaard gives a voluntaristic justification of morality: what there is reason (including moral reason) to do for any particular individual is derived from

this person's own reflectively endorsed desires. The obvious problem facing someone who wants to trace the foundation of morality to the agent's own will is relativism. For if what there is moral reason to do is ultimately a matter of which desires an agent can reflectively endorse, then, given an expectable difference amongst people, what there is moral reason to do will seem to vary in an unappealing way with those desires. Korsgaard tries to answer this threat in the course of her argument, by arguing that if we conceive of ourselves as human beings, we will naturally be lead to will the morally right thing. What interests me in this argument is the belief that one can move such a voluntaristic basis – from an agent having a reason to do something if and only if he can endorse a desire for it – to there being reasons for action which may not be in keeping with his desires, which moral reasons in particular certainly need not be. In this section I focus on the relationship between Korsgaard's project and her voluntarism, and argue that it cannot succeed on such a foundation.

According to Korsgaard, the normative question has its basis in the fact that human beings, unlike other creatures, are self-conscious. This fact implies that we cannot merely be passively at the mercy of our desires or perceptions, in the sense that we cannot just act or believe – we need reasons for doing so. This is because "our impulses must be able to withstand reflective scrutiny." (p. 93) We, that is, our "thinking selves", can monitor our impulses to act and to believe and ask whether they really are reasons for doing what our "acting selves" are disposing us to.

But this reflectivity or self-consciousness, presents us with a problem: we must now decide whether there is in fact a reason to act. But how do we determine whether a given impulse is in fact a reason? How do I decide, for instance, whether someone's insulting behaviour towards me is a reason to retaliate against the wrong done by giving the provocateur a proper beating, as I am presently impelled to? According to

Korsgaard, the solution to the problem of when an impulse – be it to believe or to act – constitutes a reason, is a matter of being able to reflectively endorse the impulse:

"if I decide that my desire is a reason to act, I must decide that on reflection I endorse that desire." (p. 97)

An impulse is transformed from the status of being a mere possible reason to act, to that of being a genuine reason, by being endorsed by the reflective self. Consequently, if I upon reflection decide that I can endorse my impulse to hit the provocateur repeatedly, then I have a reason for doing so. But by virtue of what do I resolve that this is the case?

Korsgaard's answer is that a person determines which impulses to endorse, and hence which reasons he has, by finding out whether the actions which they are reasons for could be regarded by the agent as expressive of his own values. In Korsgaardian terms, it is the agent's 'practical identity' (a concept to which I shall return shortly) which gives rise to reasons in the agent. Similarly, it gives rise to obligations:

"Your reasons express your identity, your nature; your obligations spring from what that identity forbids." (p. 101)

So whenever I can see an action as an expression of my practical identity, the impulse to perform the action will constitute a reason for me to do it, and if I cannot perform an action without violating my practical identity, then it will be obligatory for me to perform it. To exemplify: if I cannot see giving this provocateur a good beating as an expression of my practical identity then I have a reason as well as an obligation not to do so. I shall refer to this principle – i.e. the principle that if and only if an action is expressive of a given agent's practical identity, then he has a reason to do it – as *the*

Principle of Voluntarism (henceforth "PV"), since it bases practical reasons in what the agent can will without violating his own integrity.

At least to someone who believes that there are certain moral obligations which apply to agents categorically – regardless of their particular aims and interests – it seems relevant to ask whether, on Korsgaard's construal, there are things which morally agents must do. Since the answer, to Korsgaard, always turns on particular agents' conceptions of their practical identities, she does not seem able to make room for the useful distinction between objective and subjective reasons.[16] As long as the agent is fully informed and able to process the information rationally, the consideration an agent takes to support a course of action will actually be a reason for it; it is not possible that the agent might have failed to see that some other consideration supported a given course of action. For *per definition*, if the agent does not take it to be a reason for action, then it is no reason. This in turn means that Korsgaard's endorsement of PV invokes a risk of relativism, i.e. it being possible that conflicting moral judgements are both correct, but she embraces this, and believes she can dispatch it in the course of the argument.

Korsgaard does mention the Kantian argument that we are subject to certain categorical demands stemming from our practical rationality, namely those dictated by the categorical imperative. However, Korsgaard thinks there is a distinction between the categorical imperative which tells us which maxims one can will to be laws on the one hand, and the moral law – or what Kant calls the Kingdom of Ends – on the other. Moreover, she admits that the "argument that shows that we are bound by the categorical imperative does not show that we are bound by the moral law." (p. 100) Hence, since Korsgaard's argument merely shows that one has reason to act on maxims which one can will to be laws, and since such maxims may not be a part of

[16] "An obligation always takes the form of a reaction against a threat of loss of identity." (p. 102)

the moral law, whether one has reason to act in accordance with the moral law depends entirely on what practical identity one adopts.[17]

Korsgaard explains the concept of practical identity as

"a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking." (p. 101)

Accordingly, what PV states one has reason to do and what one is under an obligation to do are relative to whatever description the agent can give of himself, as long as it is one within which he values himself.

Our practical identities, Korsgaard concedes, may change as a result of the change in sympathies, commitments, and interests, which ensues partly from change in one's circumstances, partly from the knowledge one acquires as one goes through life. Accordingly, different elements of our practical identities may be shed as we lose interest in them, or come to find them flawed in some way. I may lose my interest in playing the violin and hence no longer regard being a violinist a part of my practical identity, and subsequently no longer have any reason to practise my play. Likewise, I may upon rational reflection come to view my aim of being the coolest guy on the street as shallow or frivolous, and hence lose the description 'the coolest guy on the street' as a conception I have of my own practical identity, and therefore in turn no longer have any reason to buy expensive, stylish clothes, because it would no longer

[17] This in turn is why I have called Korsgaard's approach 'the appeal to integrity' and not 'the appeal to autonomy', although she herself thinks her approach will show that "autonomy is the source of obligation" (p. 91). For clearly the notion of autonomy within which she is operating is different from Kant's, according to which what there is moral reason to do is not dependent on our particular conceptions of ourselves (that we, as it turns out, construct ourselves). For a good explanation of this difference and the importance of the distinction between Korsgaardian and Kantian autonomy, see G. A. Cohen's commentary, pp. 167-177.

be a treat to my identity if I did not look elegant. In this way, the components of my practical identity are subject to prioritization and rejection according to change in my desires, etc.

Up until now, Korsgaard has argued that human beings must, because of their self-consciousness, act for reasons, and that members of that species' reasons and obligations are founded in their practical identities, where a practical identity is a description within which an individual values himself and finds his life worthwhile. Thus, the importance of maintaining some integrity is the source of our reasons to act: they are those impulses endorsement of which are consistent with our conception of ourselves in the sense that the action which would result from acting on it could sensibly be attributed to someone who has that conception of himself.

Moreover, our obligations are those acts we must perform in order to keep the identity of someone who has endorsed whatever particular values one considers to be a part of one's conception of oneself. This voluntarism clearly makes reasons and obligations relative to what a particular agent thinks of himself, and so far at least, Korsgaard's view on the source of normativity is entirely compatible with the Williamsonian one outlined in the introduction.

However, the next step in the argument goes some way towards eliminating this relativism – at least according to Korsgaard. For she believes that there is one feature common to all agents, in terms of which they cannot but value themselves: their humanity. I now turn my attention to this second stage in the argument.

Korsgaard argues that, although one's practical identity is perpetually changing, there is something that remains constant, namely that I have to be governed by some conception of my practical identity or other. But how can that be if all reasons according to PV spring from my practical identity? Korsgaard's answer is that the reason we have for having reasons at all, that is to say, that the reason we have for

treating our particular practical identities as yielding *reasons* in the first place, is a consequence of our valuing our humanity itself:

"unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another – and with it, your grip on yourself as having any reason to live and act at all. But *this* reason for conforming to your particular practical identities is not a reason that *springs from* one of those particular practical identities. It is a reason that springs simply from your humanity itself, from your identity simply as *a human being*, a reflective animal who needs reasons to act and to live. And so it is a reason you have only if you treat your humanity as a practical, normative, form of identity, that is, if you value yourself as a human being." (p. 120-1)

So an agent must act in accordance with his practical identity, because of the value he places on his humanity. But why then must he value his humanity? Because otherwise he would not have any reason to act or to live at all, the Korsgaardian reply seems to be. Hence you must value your own humanity – "your identity as a human being, a reflective animal who needs reasons to act and to live" – if you are to have any reason to act and to live at all. Your integrity as a human being simply presupposes integrity to your humanity. Let me consider the cogency of this inference.

The argument proceeds from (1) the necessity of having some conception of our own practical identities (a description under which we value ourselves), if we are to have reason to live and act at all, to (2) the necessity of *valuing* our humanity if we are to have any reason to live and act at all. The transition from valuing one's own particular practical identity to valuing one's humanity seems to hinge on the premise – explicit in the above quotation – that our reason for valuing our particular practical identities springs from the value we place on our identity as human beings (i.e. our humanity). Hence we must value our humanity, since otherwise our practical identity could not, given PV, have given us reasons to act in the first place. On this

interpretation, our humanity may be regarded as a kind of 'meta-practical identity', which we must endorse if we are to take seriously the reasons stemming from our practical identities. We have to do that, the argument goes, since otherwise we would have no reason to act at all.

"Since you are human you *must* take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must value your own humanity if you are to act at all." (p. 123)

But must I value my humanity because it is the source of the necessity of my acting for reasons? Certainly it does not generally follow from the fact that E is a result of C, and that I take E to be a reason for acting that I must value C if C is responsible for E being a reason for acting. However things might be different in the case of humanity. Assuming that my humanity is the source of my having reasons for acting - why should I value it because it is ultimately responsible for me having reasons for acting? Because I am grateful for it? That does not seem to make sense, since one can only be grateful to someone for having been given something one would not have got had it not been for that someone. But there is little reason to believe that there is 'someone' responsible for one's being born human. Might it mean then, that one is glad that one was born a human, and not an intellectually inferior creature? This suggestion runs into the objection that people still take themselves to have reasons although they are unhappy about being human. Someone might have preferred to have been someone's pet turtle, because he prefers a comfortable existence to an enlightened one, with its attendant intransigent demand for justification and understanding. Another person might, as Korsgaard mentions (at p. 15), not value belonging to a species that produced the Nazis.

It seems however that such people could still reasonably take themselves to be subject to obligations – they could not shun their duties because they did not value themselves *qua* human beings. Even if I find my own life utterly unimportant and worthless, I still have an obligation to fulfil my promises: I cannot dissolve my promise by referring to the fact that I find my life worthless.

What is more, it seems possible to say a person can value keeping his promises without valuing being alive, indeed without valuing being a human being, "a reflective animal who needs reasons to act and to live." (p. 121) For even if one does not regard one's own life as valuable, one may still regard others' lives as valuable, and hence do what one can to contribute to their well-being, which one can do by keeping one's promises to them. Likewise, someone may simply consider it indecent not to fulfil his promises, regardless of whether it benefited anyone, and still not value being human.

Korsgaard's explanation as to why we have to value our own humanity also seems to conflict with her voluntarism. For if someone does not regard his own humanity as valuable, how can it be a part of his practical identity, and hence be a basis of reasons for action? Someone who is utterly immoral may have only immoral reasons to act, because he would be able reflectively to endorse only immoral impulses since only they are expressive of his practical identity.

In his commentary on *The Sources of Normativity*, G. A. Cohen describes such a person: the Mafioso who

> "does not believe in doing unto others as you would have them do unto you: in relieving suffering just because it is suffering, in keeping promises because they are promises, in telling the truth because it is the truth, and so on. Instead, he lives by a code of strength and honour that matters as much to him as some of the principles I said he disbelieves in matter to most of us. And when he has to do some hideous thing that goes against his inclinations, and he is tempted

to fly, he steels himself and we can say of him as much as of us, with the same exaggeration or lack of it, that he steels himself on pain of risking a loss of identity." (p. 183)

Cohen's Mafioso example is meant to square with Korsgaard's view of the basis of obligation, which, as he rightly says is "content-neutral" (p. 184); it does not take account of the content of a given individual's particular practical identity. The Mafioso struggles to overcome his tendency to flee, just like moral people battle with their inclinations to do bad things. Each type of character fulfils Korsgaard's criterion of being obligated.

Now, Korsgaard, as we have seen, has a story to tell about what explains the normativity of the actions one can reflectively endorse. She summarizes the explanation in her *Reply* as

"the necessity of living up to some of our roles, of maintaining some sort of integrity as human beings. It is the value we place on our humanity that stands behind our other roles and imparts normativity to them. And if my other arguments work, that means we are committed to valuing the humanity of others as well." (p. 256)

We have yet to look at what those other arguments are and whether they work – I turn to this in section 1.3. For now let us note that Korsgaard holds that this explanation of normativity is not what *justifies* actions:

"I have offered an explanation of the existence of moral obligations which, I claim, should lead you to endorse those obligations, unless you are prepared to be a complete sceptic about reasons and values. But (assuming that the argument is successful) what does the work here, your reflective endorsement or the explanation itself? Now as the caveat about avoiding scepticism shows, I must say that it is the endorsement that does the work, since I am prepared to agree that if human beings decided that human life was worthless then it would be worthless." (p. 254)

Since it is the reflective endorsement that does the justificatory work, it follows that the Mafioso has an obligation to, say, eliminate competition in his line of work. For to Korsgaard, as I mentioned, "[a]n obligation always takes the form of a reaction against a threat of identity." (p. 102)

I conclude therefore that since Korsgaard's voluntarism trumps the explanation of why we have obligations, it seems impossible to justify moral conduct on Korsgaardian grounds. Although it may be correct that the explanation of the existence of moral obligations shows that everybody, by virtue of their humanity, has moral obligations, one would still have to endorse one's humanity in order for one to be justified in acting on those obligations. Thus if someone did not endorse his humanity, he would have none of the obligations that according to Korsgaard's argument stem from valuing one's humanity.

Moreover, even if we grant that PV is true, such that there is reason to perform an action if it is expressive of our practical identity, it still does not follow that if D, a description under which we value ourselves, is ultimately responsible for our having reasons to act, we must also value the alleged origin of our ability to value D, for it is, in Korsgaard's terms, "the endorsement that does the work", and not the explanation. Therefore unless we endorse our humanity we have no reason to heed whatever obligations might have flowed from so doing.

However, this seems, as I pointed out above, a completely counterintuitive implication of Korsgaard's view. It seems that people might quite reasonably take themselves to have moral obligations, or more generally, reasons to act although they do not value their own lives (if this indeed is what valuing one's humanity is supposed to mean). If somebody does not value his own life at all, he may still take himself to have duties towards others. This seems to suggest that there is no necessary connection between having reasons to act and valuing one's own humanity.

Hence, although (1) may be true, such that agency requires that we have a practical identity, it does not follow from this that we must value our humanity. For first of all one can take oneself to have practical reasons despite the fact that one does not value one's humanity, and secondly, since according to Korsgaard's theory, voluntarism is the real source of normativity, those obligations that follow from valuing one's humanity, would fail to give one reasons to act unless one reflectively endorsed it.

Let me continue by considering the third and fourth steps – the ones which, on the basis of the framework outlined in the introduction appears to be the most controversial of them all. This is because they involve an attempt to move from agent-neutrality to agent-relativity, merely with PV as a starting point – a principle which ultimately seems to be an expression of a kind of internalism about reasons akin to the Williamsonian one outlined in the introduction.

Assuming for the sake of the argument that Korsgaard's inferences work thus far, we now have reason to believe that when we act for reasons we are committed to the value of our own humanity. The third stage goes on to argue that valuing our own humanity requires valuing that of others as well, whilst the fourth and last stage argues that valuing humanity as such conveys moral obligations on us:

"valuing ourselves as human beings involves valuing others that way as well, and carries with it moral obligations." (p. 121)

Surely an astonishing inference. I want, first, to explain briefly what I take to be Korsgaard's rationale for the two claims she is making – i.e. one the one hand (3), that valuing one's own humanity requires valuing others' humanity as well, and (4), that this conveys on us moral obligations, on the other. Then I will continue, in the following section, with a more thorough assessment of (3), the step which I take to be the most crucial one.

Korsgaard goes about arguing for (3) by attempting to bridge the gap between agent-relative and agent-neutral reasons, or in Korsgaard's preferred equivalent terms, private and public reasons.[18]

The first thing worth considering is how this will get us to (3), which is an evaluative, and not a normative statement. How, if we assume that the argument from agent-relative to agent-neutral *reasons* goes through, do we get from agent-neutral reasons to agent-neutral *values*? It seems that the thought is that just as I have reason to value my own humanity because it is the ultimate source of my having reason to do what my practical identity dictates, so, analogously, I have reason to value everyone else's humanity insofar as *it* is the source of reasons for action. Reasons and values, according to Korsgaard's general framework, go hand in hand: if there is reason to do something, then there is reason to value whatever is the source of those reasons being reasons. Hence, if it can be shown that I have an agent-neutral reason to be concerned about humanity wherever it is found, that would also entail that I have reason to value humanity wherever it is found, that is, that humanity is agent-neutrally valuable.

As regards the last step in Korsgaard's argument, from (3) to (4), it can be summarized as follows: if indeed there is no sound distinction between agent-relative and agent-neutral reasons, then neither, as we have just seen, is there a sound distinction between agent-relative and agent-neutral values because practical reasons are nested in the value of things. If, then, agent-neutral values are part of my practical identity I have reasons stemming from agent-neutral values. Because my obligations in general spring from what my identity forbids, I have moral obligations – obligations engendered by the value of others – since agent-neutral values necessarily are part of my practical identity.

[18] Cf. p. 133n3; 221.

I shall not discuss the last inference here since it is largely irrelevant to my purposes. Instead I will focus, in the subsequent section, on what is more important for the general theme of this thesis, namely the crucial argument against a sound distinction between agent-relative and agent-neutral reason.

Having now distinguished and explained the main steps in Korsgaard's argument, let me summarize it as follows:

(1) You must have a practical identity, i.e. a description under which you value yourself.
(2) If you have a practical identity then you must value your humanity.
(3) If you value your own humanity then you must also value humanity as such.
(4) If you value humanity as such, you must have moral obligations.

1.3 FROM AGENT-RELATIVE TO AGENT-NEUTRAL REASONS

In this section I want to have a closer look at the crucial inference signalled by (3): the inference from agent-relative to agent-neutral reasons. Korsgaard describes the project as that of showing that agent-relative *reasons* are merely incidentally agent-relative:

> "To act on a reason is already, essentially, to act on a consideration whose normative force may be shared with others. Once that is in place, it will be easy to show how we can get someone who acknowledges the value of his own humanity to see that he has moral obligations." (p. 136)

But surely, the fact that the normative force of a reason *may* be shared with others does not go very far towards showing that it *will* or *should* be. The latter conclusion is needed in order to show that people have moral obligations: Stating that reasons are share*able* does not certify the conclusion that they are in fact shared, nor that they ought to be. I can recognize your reasons for action as having motivating force for

you, and you can recognize mine as having motivating force for me, but that we can each understand the motivating force of each others' agent-relative reasons – that they are in that sense shareable – does not make us share them or show that we ought to. For example, I may realize that the fact that my friend is an enthusiastic birdwatcher gives him a reason to drive hundreds of miles to a place where he has reason to believe a rare and exotic specimen has been spotted. But that I realize, and perhaps even understand, that being an enthusiastic birdwatcher gives the birdwatcher a reason to travel a great distance in an attempt to spot a bird, does not give me a reason to do so, since I do not feel the same way about birds. Neither does it show that I ought to feel the same way about birds. Clearly there is a distinction here between shareable and shared.

Korsgaard addresses this concern, and her answer involves something which I shall call *the Intrusion Thesis*, namely the claim that we always have the ability to "intrude" (p. 136; 139) ourselves into each others consciousness just by uttering certain words. By doing so we make our reasons know to someone else, whereby we obligate each other. Korsgaard appears to believe that, as long as I am able to communicate the meaning of a sentence expressing my reason for acting then insofar as I make you grasp my reason, the mere uttering of these words will be sufficient to force any listener to be motivated to do what I take myself to have reason to do. This, I take it, is Korsgaard's idea of how the vacuum between agent-relative and agent-neutral reasons is to be filled up.

Hence the argument for (3) involves these two steps: firstly, reasons can be shared, that is, they are share*able*, and hence not necessarily private. Secondly, because of the truth of the Intrusion Thesis, they will be shared whenever we become aware of the reasons each of us have. This in turn would mean that all agent-relative reasons are potentially agent-neutral reasons, which in its turn, as I explained at the end of the

previous section, would mean that humanity as such has agent-neutral value. Let me consider the cogency of the two steps in the argument in turn.

In respect of the first step, Korsgaard begins by identifying what she takes to be the culprit responsible for the allegedly mistaken belief that reasons cannot be shared: the idea that the activity of reflection is an un-shareable activity.

> "People suppose that practical reasons are private because they suppose that reflection is a private activity. And they suppose that, in turn, because they believe in the privacy of consciousness." (p. 136)

So had people not believed in the privacy of consciousness, they would not have believed in the privacy of reflection, and had they not believed that in turn, they would not have believed that practical reasons are private. According to this clause, in order to establish that people have reason to believe that practical reasons are shareable, we need to show that the belief that consciousness is private is wrong.

What does it mean that consciousness is private? We get some clue as to what Korsgaard means from the fact that she goes on to argue against the privacy of consciousness, by an invocation of Wittgenstein's private language argument, which is widely considered to be an argument to the effect that there can be no such thing as a private language. Hence, the thesis that consciousness is private involves at least the claim that it is possible to have private languages, i.e. words and sentences the meaning of which are debarred from being communicated or shared. This is a much stronger sense of the privacy of consciousness than the common-sense version of the idea which merely understands by the privacy of consciousness something like the proposition that people have privileged access to their own mental events, and that they can choose to keep the content of their mental activities from other people.

Korsgaard must, however, since she goes on to argue against the privacy of consciousness by invoking Wittgenstein's private language argument, understand

the term as signifying the acceptance of the belief that the meaning of certain mental terms cannot be shared, in the sense that they are incommunicable to others.

Just as Korsgaard believes that the private language argument shows that meanings cannot be in principle incommunicable, so she believes that a private *reasons* argument can be constructed along similar lines. According to Korsgaard's interpretation, the reason why there cannot be a private language is that it is inconsistent with the normativity of meaning, viz. the claim that "to say that X means Y is to say that one ought to take X for Y" (p. 137). This, Korsgaard believes, "requires two, a legislator to lay it down that one must take X for Y, and a citizen to obey" (ibid), for had there not been these two, it would not have been possible to be mistaken about the meaning of term, which would be absurd, because then it would have no meaning at all. There can be no terms with private meanings, because one could not be erroneous about the truth-conditions of such terms.

Korsgaard thinks a parallel argument can make a case against the existence of private reasons:

> "reasons are relational because reason is a normative notion: to say that R is a reason for A is to say that one should do A because of R; and this requires two, a legislator to lay it down, and a citizen to obey." (p. 137-8)

Nonetheless, this argument does not establish that there cannot be purely private or agent-relative reasons. For although Korsgaard seems to intend the argument to show this, the legislator and the citizen of whom she speaks are regarded by her as elements of the same person's reflective consciousness, viz. "the thinking self and the acting self" (p. 138), i.e. the reflective and executive parts of human agency. But since both of these elements belong to one and the same individual we have not yet moved beyond agent relativity: it is still the same individual that Korsgaard is referring to, so she is not claiming that there have to be more individuals in order for something

to be a reason (like there has to be more than one individual in order for words to have meanings).

So far then Korsgaard has argued that reasons are shareable. She continues on this assumption, and goes on to argue for what I have called the Intrusion Thesis, that is, that we can give each other practical reasons whenever we want to – reasons which would then weigh with the other, and carry genuine motivating force. Korsgaard argues for this claim, by bringing forth certain examples where this seems to be happening. For instance, she believes that if someone calls out another's name, then he has by that very act, given the other a reason to stop:

"that I have given you a reason is clear from the fact that, in ordinary circumstances, you will feel like giving me one back. 'Sorry, I must run, I'm late for an appointment.' We all know that reasons must be met with reasons, and that is why we are always exchanging them." (p. 140)

To Korsgaard this is taken as a proof of the fact that we take others' reasons into account by default:

"the reasons of others have something like the same standing with us as our own desires and impulses do. We do not seem to need a reason to take the reasons of others into account. We seem to need a reason not to." (p. 140-1)

The claim seems to be that all reasons are agent-neutral: there is no sound distinction between what I have reason to do and what you have; there are, as it were, only reasons. And the problem of how we become motivated to act on those reasons is not that of how we come to be motivated by them, but rather how we become aware of the existence of the reason. Once someone has made it clear to us that there is a reason for us to stop, we automatically see its reason-giving force and are motivated to stop. This, supposedly, is revealed by the fact that if we do not stop, we need to

state, or at least have at hand, a reason why we should not which is capable of overriding this reason. So, according to Korsgaard, we do in fact incorporate the reason we are given into our deliberations. Hence reasons are supposed to be inherently agent-neutral.

First of all, it seems wrong to base a normative conclusion about how we have reason to act on the basis of descriptive premises. It may well be that people act in this and that way, and that this in turn indicates that people treat the desires and impulses of others as giving them reasons to act, but this could never in and of itself justify them in doing so. Even if all of humanity conducted themselves in that way, it would never show that they have an agent-neutral reason to do so. For that conclusion to be warranted one would need to show that there is something objectively recommending it.

Secondly, however, the claim that we do in fact incorporate the reasons of others seems to be flat out wrong. For example, as G. A. Cohen notes in his commentary (p. 176), if someone calls me he might just as well have given me a reason to speed up. Korsgaard admits that the rational egoist would object to her argument that the fact that we seem to need reasons not to take other's reasons into account justifies the conclusion that reasons are agent-neutral or public, on the ground that

> "this does not establish that other people's reasons are reasons for me. I [= Korsgaard] am merely describing a deep psychological fact – that human beings are very susceptible to one another's pressure. But nothing I have said so far shows that we really have to treat the demands of others as *reasons*." (p. 141)

Thus, the egoist seems to be objecting that even if we realize that someone else has an *agent-relative* reason to want us to stop by calling out our names this does not necessarily show that we have a reason to stop. Korsgaard needs to address the concern that even if we recognize that some else has an agent-relative reason to want

something to happen this fact cannot in and of itself give us reason to want it as well. Something extra seems to be required.

Korsgaard tries to address this issue by describing a case in which two people discuss when a meeting is to be held – a case in which the two interlocutors are already obviously motivated to find a solution. Korsgaard argues that egoism is to be rejected, because it is unable to account for the phenomenology of such cases. Here is the case which Korsgaard takes to pose difficulties for egoism:

"A student comes to your office door and says: 'I need to talk to you. Are you free now?' and you say 'No, I've got to finish this letter right now, and then I've got to go home. Could you possibly come around tomorrow, say about three?' And your student says 'Yes, that will be fine. I'll see you tomorrow at three then.'" (p. 141)

Korsgaard takes this to be a case of two persons "reasoning together" to arrive at a decision about what to do. However, says Korsgaard, if egoism is true, then this is not what is happening. Instead, the each of the two parties

"backs into the privacy of his practical consciousness, reviews his own reasons, comes up with a decision, and then re-emerges to announce the result to each other. And the process stops when the results happen to coincide, and the agents know it, because of the announcements they have made to each other." (p. 141)

But why is this description not fitting for the case at hand? Argumentation to the effect that it is not is entirely deficient. This lecturer/student-example of Korsgaard's is entirely consistent with the egoist's explanation, as long as no reason has been given as to why the lecturer and the student are motivated merely by each other uttering certain words. The egoist has not been prevented from explaining the observation that the lecturer and the student are motivated by a desire for a common goal (arranging the meeting), by making reference to the fact that they are already

motivated by other things. Thus the lecturer might realize that she has an obligation to supervise students as a result of her having signed a contract when she took the job, regarding it nevertheless as a necessary evil, required of her in order for her to remain in an institution where she can pursue her *real* interest, research. Similarly, the student may be motivated primarily by his desire to learn, and he may have no choice but to see his supervisor when she can find the time. Hence neither party, on this quite plausible, albeit egoistic, construal, are motivated by the other's reason. But – and this is the point – they are "reasoning together" nevertheless. Therefore Korsgaard cannot use examples such as these to answer the egoist's objection that people need a prior motivation to adopt others' agent-relative reason. If this is correct, then reasons will not be shared merely by bringing the other to see the considerations which you took to be reasons, and the Intrusion Thesis will be incorrect.

It seems that we have been given no reason as to why we should not endorse egoism as the most likely explanation of what is going on in the lecturer/student-example. Exactly the same point seems to apply to the second of Korsgaard's examples, the tormenter/target-example:

> "Suppose that we are strangers and that you are tormenting me, and suppose that I call upon you to *stop*. I say: 'How would you like it if someone did that to you?' And now you cannot proceed as you did before. Oh, you can proceed all right, but not just as you did before. For I have obligated you to stop." (p. 142-3)

In this attempt to derive moral obligation from a mere speech act, Korsgaard relies on the Intrusion Thesis, for she believes that the motivation to stop comes about merely by making the tormenter see that he would have resented it if he had been the target, and furthermore that the mere speech act will yield this result. In short: by

uttering 'How would you like it if someone did that to you?', the target confers on the tormenter an obligation to stop.

There is an obvious objection to this though. If the tormenter is an egoist, which Korsgaard's setup surely does not exclude him from being, he is likely to reply to her question by remarking that he would not like it at all if someone did that to him, since his pain is something undesirable *for him*, just as the target's pain is undesirable for her. So the tormenter has a reason not to experience pain just as the target has. But the rational egoistic tormenter will deny that the target's pain is a reason for her, just as his pain is not a reason for her. As pointed out earlier this is an entirely consistent position, and there is nothing in the logic of the tormenter/target example that should persuade us to think otherwise.

Korsgaard believes this is wrong, however. For if I do not come to be motivated by your agent-relative reason when you ask me *qua* tormenter how I would like to go through what you are going through, then

"I would have to hear your words as mere noise, not as intelligible speech. And it is impossible to hear the words of a language you know as mere noise." (p. 143)

But again, this is not a satisfactory reply, for as Thomas Nagel very neatly points out in his commentary:

"The invocation of Wittgenstein doesn't help, because egoism doesn't violate publicity." (p. 208)

This is exactly the point: the egoist may understand that the severity of the pain is a reason *for the target* not to undergo it, but this does not entail that *he* is moved by this consideration. Korsgaard seems to base her argument on a false dichotomy between agent-relative reasons, which are essentially un-sharable on the one hand, and agent-neutral reasons which are shareable on the other. This dichotomy is false because

agent-relative reasons *are* shareable: the egoist *can*, as Nagel points out, share someone else's reasons, in the sense of understanding their force for this individual. The objection that we are either moved by someone else's reasons, or fail to understand the reason (*for her*) entirely is mistaken, and hence the gap between agent-relativity and agent-neutrality remains intact.

Perhaps Korsgaard's idea of the inherently motivating force of words in the tormenter/target-example draws some support from its resemblance to the lecturer/student-example in the following way: had the lecturer said to the student that she did not have time to see him at all, the student might have lectured the lecturer by saying that she in fact has an obligation to do so, because it is part of her job – part of her practical identity, we might say. She values doing the research, and since supervising, though undesirable in itself, is a necessary activity in order for her to keep a position in which she can do research, she has to supervise in order to do what she finds truly worthwhile.

Now to many people, being morally good is likewise a part of their practical identities; they consider fulfilling their moral obligations a part of what makes life valuable. Hence, if they do not conduct themselves accordingly, we can normally bring them to be motivated to do the right thing by pointing out to them that it is indeed the morally right thing. But that move is not possible when confronted with the rational egoist, for he just denies that there is reason to do the morally right thing irrespective of whether it constitutes or brings about a benefit to him. Therefore, making him realize that he is actually doing something morally repugnant will not have the same effect. The two examples seem to run parallel only if we assume that the lecturer is motivated to do her job and the tormenter to be morally good. But given the conclusion from section 1.2, Korsgaard has to show why someone who is presently without any motive to be morally good has reason to be motivated by moral considerations.

Her argument as we have seen does not establish this. It is as if she refuses to recognize that there is a difference between listening to, and understanding someone stating his reason for acting on the one hand, and being motivated by such considerations on the other. Although Korsgaard tries to break down this dichotomy by quoting examples where this appears to be the case, she does not succeed in dispatching the concern that the examples merely draw support for her thesis because of the intuitive appeal of the thesis itself. In other words the conclusion she draws from the discussion of her examples presupposes what she is trying to argue, namely that understanding another's reason is inherently motivating. For the examples clearly seem rigged so that one party in the exchange of reasons already possesses a prior motivation either to help or to avoid suffering. This cannot be presupposed if one intends to argue that understanding the reason in and of itself is responsible for the motivation. In effect, Korsgaard is unable to address the challenge posed by the egoist who understands the reason *qua* agent-relative, but has no motivation whatsoever. In other words, nothing Korsgaard has put into play has shown that the distinction between agent-relative and agent-neutral reasons is undermined. I therefore conclude that the crucial third premise is unwarranted.

1.4 CONCLUSION

In this chapter I have pointed out that Korsgaard's attempt to argue someone into accepting moral demands fails. In section 1.1 I pointed out that Korsgaard was indeed committed to the claim that all agents could be given a satisfying answer when they posed the normative question. In section 1.2, I summarized and explained the steps in this answer, one which started from the necessity of human beings acting for reasons, and went on to our valuing our practical identities, and from our valuing our practical identities to our valuing our humanity, and from there to our valuing humanity at large, and finally to our having moral obligations.

It was shown how several of these inferences were dubious. In particular, I pointed out that since Korsgaard believes that the ultimate source of normativity is conceptions of oneself within which one finds one's own life valuable – conceptions which one may choose at random – it becomes difficult to see how moral reasons can apply to an agent irrespective of whether 'being moral' is a conception under which this or that particular agent values himself, which he obviously need not.

My main line of objection, however, focused on the possibility of the inference from valuing one's own humanity to valuing humanity at large, and the rejection of the distinction between agent-relative and agent-neutral reasons that it carries with it. Korsgaard's argument to the effect that the agent-relative/agent-neutral gap could be bridged was shown to rely on something I called the Intrusion Thesis – essentially, that an agent could intrude himself into the consciousness of another person merely by uttering the consideration which constituted his reason. Not, however, merely in such a way as to make the other aware of the meaning of the sentence when he utters it, but also in the sense of making him motivated by that consideration himself.

Korsgaard's Intrusion Thesis was shown to be mistaken. Her argument relied on some examples which would only succeed in establishing the proposed conclusion (that there is no sound distinction between agent-relativity and agent-neutrality), if we presupposed that the person, who allegedly became motivated by the uttering of the reason, was already in some way motivated by considerations such as those constituting the speaker's reason. This shows that the reason, although it can be communicated to others, is not inherently reason-giving. Hence Korsgaard's argument does not succeed in arguing someone into becoming moral, for if he is motivated merely by agent-relative reasons, then he ought not solely by listening to Korsgaard's argument, become motivated by agent-neutral ones.

It is also important to remember that even if the Intrusion Thesis had been correct, it could not have established the normativity of agent-neutrality, since one cannot get

from facts about what agents as a matter of fact act are like– in this case that they are sensitive to each others' desires and impulses, and treat them as reasons to act – to conclusions about what they ought to do. The normative conclusion would have to be based on some considerations about whether there is anything to recommend so acting, not descriptive observations about this or that person coming to think that there is reason for him to do something. As I pointed out in section 1.1 Korsgaard herself recognizes that a mere explanation of moral practices could not amount to a justification for those practices. Thus there is at least room for a concern that she might contradict her own axiom here.

Although Korsgaard's attempt to show that there is reason for someone external to morality to be moral ultimately fails, I believe she provides a good account of the *phenomenology* of obligation. To most of us morality is something we value – we care deeply about how we treat other people, regret if we do something wrong, and apologize, and ask for forgiveness. Likewise we feel good (or at least not bad) about ourselves when we do something morally right, feel a sense of unity with and acceptance from other people, perhaps even humanity at large. Obligations often seem to spring from the conceptions under which we value ourselves, just as Korsgaard believes it does, and practical deliberation seems, as a matter of fact, to proceed in the way she outlines: we consider whether a certain act is consistent with our endorsed values.

Nonetheless, I conclude that Korsgaard fails in her attempt to argue someone into morality, for – to use Nagel's words – egoism does not violate publicity.

2. RAZ'S APPROACH: AN ARGUMENT FROM VALUE-THEORY

In the previous chapter we saw how Korsgaard tried to bridge the gap between agent-relative and agent-neutral values/reasons, by arguing for an implausible thesis according to which it is possible for people to impose the normative force of their own 'private' reasons on others, thereby making their normative force shared by others. Had that been possible, it would have established that moral reasons are inherently agent-neutral, or 'public' in Korsgaard's terminology, such that any rational agent who becomes familiar with the existence of that consideration would recognize its reason-giving force. As we saw, however, Korsgaard's argument turned on a false dichotomy between, on the one hand agent-neutral reasons which are essentially shareable, and agent-relative reasons which are essentially un-shareable on the other. Relying on this dichotomy, and an argument to the effect that moral reasons are shareable, Korsgaard concluded that moral reasons are agent-neutral reasons. This inference should be resisted, however, since we can – for the most part at least – quite easily understand others' reasons without them necessarily being agent-neutral.

Besides, Korsgaard's argument for the thesis that agent-neutral reasons are inherently shareable was shown to hinge on an implausible optimism with regard to people's ability to generate motives in others. It appeared that Korsgaard was simply begging the question: she wanted to show that the uttering of the reason brings about a motivation, but seemed to presuppose a pre-existing motivation in the examples which she cited in support of her conclusion. In addition, I pointed out that there is something fishy about trying to establish a normative conclusion on the basis of evidence about how people as a matter of fact behave.

In this chapter and the next, we will be looking at attempts that try to argue not that our practical reasoning commits us to morality, but rather that a certain common feature of our lives does. They advance arguments on the basis of the prudential

value of friendship, or more generally, relationships with intimates, and try to show that thinking about the nature of such relationships can show that there is reason to act morally. The first of these – to be discussed in this chapter – argues that there are certain value-theoretical considerations about friendships which commit us to acting morally, whereas the second – to which I turn in the next – argues that there is reason to be moral for prudential reasons because interacting morally with other people is a way of extending our own interests into the future.

In this chapter I want to assess the value-theoretical approach to the normative question encountered in Joseph Raz's 'The Amoralist'. It is a part of Raz's general agenda that the distinction between moral reasons and prudential reasons is misconceived – that in some sense there are only practical considerations with different weights. In a series of articles, of which I focus here on one in particular, he argues that there is no fundamental difference between the normativity of prudential and moral considerations in people's practical deliberations:

"as I see it, it is true that when we deliberate we consider which reasons are most pressing in a way which transcends and defies the common division of practical thought into moral and self-interested (and other) considerations." (p. 306) [19]

Raz bases this view on a conception of agency and normativity – *the Classical Conception* – which maintains that,

"the central type of human action is intentional action; that intentional action is action for a reason; and that reasons are facts in virtue of which those actions are good in some respect and to some degree." (p. 23)[20]

[19] See "On the Moral Point of View", "The Amoralist", and "The Central Conflict: Morality and Self-Interest", chs. 11, 12, and 13 in Raz (1999). Unless I indicate otherwise, all parenthetical page references in this chapter will be to Raz (1999).

However, although it may be the case that people normally do not distinguish between prudential and moral reasons when they deliberate what to do – that they, as it were, do not consider in what *respect* their actions are good, and merely act because, on balance, there is most reason to perform this act – it certainly need not follow that these kinds of reasons cannot be distinguished.

Moral reasons include agent-neutral ones, and the fact that it might be an adequate description of most people that they recognize and act according to such reasons without singling them out from prudential ones does not show that anyone has reason to conduct himself in accordance with this convention, let alone that one is irrational if one does not. One can quite consistently believe in the Classical Conception's idea that reasons are nested in the value of things, but not believe in the existence of agent-neutral values, and therefore deny that there are agent-neutral reasons to act.

Besides the Classical Conception, Raz's approach builds on an objective account of well-being, according to which

"one's well-being depends on success in worthwhile and wholeheartedly engaged-in goals and relationships." (p. 322)

So assuming that moral goals are valuable goals, doing the morally good thing can contribute to one's well-being. In "On the Moral Point of View" Raz argues for the conclusion that doing the morally right thing can in fact contribute to the well-being of the agent, and thus that the reasons that people have to be moral derive from their interest in their own goals, aims, or well-being. This then leads Raz to defend – in "The Central Conflict" – the Classical Conception against the twin-objection that (a)

[20] Various aspects of Raz's classical conception are discussed in Ulrike Heuer, "Raz on Values and Reasons" in Wallace, Pettit, Scheffler, and Smith (eds.) (2004).

it cannot explain the conflict often felt when moral reasons propels us to act against our own self-interest, nor (b) the subsequent self-sacrifice felt when we have to give up our more self-confined pursuits in order to do what is morally right. How is such a conflict and feeling of self-sacrifice possible, if according to Raz's conception of well-being, they can clearly contribute to one's well-being, and if, further, there is no real distinction between prudential and moral reasons? According to Raz, this objection is based on the false idea that our own well-being provides us with reasons for action. Our own well-being, he argues, is an evaluative concept which, normally, does not figure in our deliberations as such, when we try to determine what to do. On the contrary, what does figure are the values which provide the reasons *directly*, as it were, unmediated by considerations as to whether acting because of reasons nested in *these* values will make our lives good.

Regrettably, I cannot offer a full assessment of these interesting papers here.[21] I will focus instead on "The Amoralist", in which Raz argues that by considering what it means to live a meaningful human life we might convince an amoral person to accept the moralist's position. Just as Korsgaard, in her way, attempts to illustrate that the agent-relative/agent-neutral distinction is bogus in practice, so Raz wants to show that insofar as a person stands in certain relationships to other people on which he places value, he is rationally committed to the agent-neutral value of people as such. This he takes to be tantamount to showing that anyone who has certain sorts of relationships with other people is committed to being moral on pain of irrationality.

In the course of this chapter I will argue that Raz's argument does not fare better than Korsgaard's, basically because it draws its force from certain flawed assumptions about value-theory: in essence it fails to appreciate fully the distinction between agent-relative and agent-neutral value. In section 2.1 I point out why Raz's

[21] For some discussion hereof, see R. Jay Wallace, "The Rightness of Acts and Goodness of Lives", in Wallace, Pettit, Scheffler, and Smith (eds.) (2004), esp. pp. 385-390.

paper should indeed be read as another attempt to argue someone into accepting moral demands. In section 2.2 I state and explain what I take to be Raz's argument, and finally, in section 2.3, I will explain why I take the argument to be mistaken.

2.1 RAZ ON THE NORMATIVE QUESTION

In 'The Amoralist' Raz intends to argue that an ordinary normal human life involves a commitment to the value of a moral life. Raz goes about arguing for this claim by first placing before us the amoralist, who, for prudential reasons, has friends, but who refuses to accept the moralist's principle, which Raz takes to be that people are valuable in themselves. Having placed such a character before us, Raz argues that this prudential value – friendship – commits the amoralist to the moralist's principle. Hence Raz clearly believes that there are values, which – although non-moral themselves – commits anyone who recognizes them as values, to morality. In other words, we are faced with another endeavour to argue that there is an intimate – indeed a necessary – connection between agent-relativity – which, as I have suggested, carries no necessary commitment to morality – and agent-neutrality, which does. Since it is part of my thesis that no such leap from agent-relativity to agent-neutrality can be made, I should like to show how the Razian argument fails. In this first section I want to explain how Raz's paper relates to the agent-relative/agent-neutral distinction and the normative question.

First of all let me note that the fact that Raz focuses on the amoralist, and not, as I have done so far, on the rational egoist, is irrelevant. Raz recognizes that his argument applies to the rational egoist himself, such that it, if successful, would show that there is reason for the rational egoist to become moral as well.[22] But what defines the amoralist?

[22] Raz calls the rational egoist the 'moral egoist', and says that he "believes that all life is of value but that only its possessor has any reason to do something about it." (p. 275) In a footnote he then says

The Razian Amoralist is someone who is external to morality in the sense explained in section 1.1, for he "does not believe in morality" (p. 273), because he "denies that persons are valuable in themselves." (p. 274) Consequently, if he had believed counterfactually that persons were valuable in themselves, he would have believed in morality.

Unfortunately, Raz does not attempt to define what being valuable in itself means. However, he does say that it is normally taken to mean that "other things being equal, an action is (morally) justified only if it can be justified taking proper account of the interests of all those whose interests it affects." (p. 275) Accordingly, Raz must take there to be an intimate connection between valuing someone in himself, taking his interests into account, and actions towards him being morally justified, such that valuing someone in himself commits one to the view that actions need to take proper account of the interests of others in order to be morally justified. Thus, according to Raz, if one believes that someone is valuable in himself, one *ipso facto* believes that there is reason to take his interests into account, and that actions towards him should be morally justified. I shall therefore understand Raz as maintaining that the moralist believes that people are valuable in themselves (or are agent-neutrally valuable) and that this means that one believes that there is reason to treat people in accordance with the demands of morality.

As I said, Raz believes that he can show that the amoralist, who denies the moralist's central belief, and who therefore believes only in the agent-relative value of people, is committed to accepting the agent-neutral value of people, and therefore to there being a reason for her to be moral.

---

that "[besides undermining the amoralist] the argument of this essay undermines the moral egoist as well." (p. 275 n. 7)

"My argument is that there are certain activities, pursuits, relationships which though non-moral themselves commit anyone who regards them as valuable to the moralist's principle." (p. 288)

Raz focuses on the prudential value 'friendship' in particular and tries to show that it commits the amoralist to the agent-neutral value of persons.

In outlining the argument, he clearly believes that the ensuing argument will succeed in eliminating the gap between the moralist and the amoralist:

"By examining the amoralist who has at his disposal the full range of goods by which his life can be enriched, and investigating the evaluative presuppositions of these goods we can – I will argue – demonstrate that there is no gulf between the moralist and the amoralist, and we can do so more securely and in a more far-reaching way than if we disregard these value-presuppositions in trying to extend the amoralist's sympathies and motivations." (p. 284)

It is remarkable, however, that this purported conclusion is far stronger than – not to say in direct contradiction with – what Raz takes his argument to establish after having put it forth:

"It is not my claim that there is a single knock-down argument which shows that the amoralist in recognizing values which can enrich his own life is committed to recognizing the value of all people. My suggestion was that there is enough in what he is committed to to advance the argument and narrow the gap between the amoralist and the moralist." (pp. 298-9)

Determining whether Raz should be understood as trying to establish the stronger claim that there is no gap between the amoralist and the moralist, or the weaker one that the gap between them can be narrowed, must hinge on a proper analysis of the argument. I pursue this in sections 2.2 and 2.3 below. However, whether or not Raz actually believes that the gap between the moralist and the amoralist can be completely eradicated or merely narrowed, the argument is still relevant to my

general thesis, according to which there is nothing in the concept of agent-relativity that, taken in isolation, could bring us to the recognition of agent-neutrality required by morality.

So far I have explained that Raz's argument pertains to a certain gap between the amoralist and the moralist, a gap which he thinks can be either be eliminated or narrowed by investigating the value-presuppositions of the prudential goods endorsed by the amoralist. I have also explained that this gap clearly relates to my thesis that one cannot leap over the gap between agent-relativity and agent-neutrality. Raz clearly believes that someone external to morality, in the sense that he is not motivated by moral considerations at all can be shown to be committed to the normativity of moral considerations.

2.2 THE ARGUMENT

Because Raz's basic idea is that someone who wants to live a prudentially good life has reason to live a moral life, he thinks he can legitimately ignore the amoralist

> "who has no concern for people, no friendships, no people he likes or is fond of, and who has no desire for such feelings, attitudes, and relationships. Such a person's well-being is drastically affected by these limitations. Not only relationships, but also all the activities which depend on them or which presuppose the appreciation of their value are denied him. This means, for example, that his ability to appreciate and benefit from literature and the arts, and from many social activities, is severely limited. There are things he can enjoy. But his life is so severely limited that […] he poses no challenge to morality. The challenge is posed by an amoralist who can have a rich and rewarding life, while denying the value of people. Such an amoralist is like us in valuing friendship and companionship. He cares, however spasmodically, for some people. Is it possible to be a consistent amoralist of this kind?" (p. 283-4)

The fact that Raz makes whether or not someone poses a challenge to morality hinge on whether or not he can live a rich and rewarding life may appear puzzling. After

all, would not anyone have reason to do the morally right thing irrespective of whether they were able to live prudentially good or bad lives while doing so?

Perhaps Raz would agree with this, but his assumption makes sense in the context of his argument, since he wants to argue someone into accepting moral demands. Hence, Raz assumes that he can ignore someone who has no concern for others, since he would clearly have a reason to, because of the objective good realized in being so concerned.

In accordance with Raz's objectivist theory of well-being, there might be reason for someone to pursue friendship although he has no present desire for it, simply because friendship is desir*able*. That friendship is desirable would in this context mean that there is something about friendship which makes it worthwhile independent of anyone's actual desires.

This view seems correct to me – there is reason for someone to desire friendship, although not a conclusive one. If, for example, an agent is too busy with other valuable pursuits, or if he does not possess the basic capacities and sensibilities required for the success of friendship, then it seems he has no reason to pursue such relationships.[23] However, that does not render friendship undesirable – there is still something counting in favour of the pursuit, maintenance, and development of friendship, although this particular agent has no such reason.

[23] Something to which Raz himself is committed because of his theory of well-being, which takes *endorsement* of the goals and activities one pursues to be a necessary condition hereof (see p. 322; cp. 1986: 291-2). Assuming that an agent cannot endorse some friendship for lack of a desire to be in it, it could not contribute to his well-being, and accordingly he would have no reason to be in it. As we shall see later, Raz also affirms that an agent has to have the capacities required to engage with some good in order for it to be good for him. For he holds that to justify that something is good for an agent one must show "(a) that the thing is good, and (b) that the agent has the ability and the opportunity to have that good." (p. 260)

But it still seems that Raz can legitimately ignore the amoralist who has no desire for friends, since he would have an objective reason to value – although not necessarily pursue – friendship.[24] That reason would stem from its objective desirability, and the amoralist would therefore be committed to whatever a commitment to the value of friendship carries with it.

So let us assume then that the amoralist has friends and that in case he has not, that he would have reason to value them. The tenets of Raz's argument means that now he has to overcome the difficulty represented by the amoralist who, in Raz's words, enjoys the benefit that friendship conveys, whilst just caring for people in a spasmodic manner, i.e. not believing that they are valuable in themselves, independently of his interest in them. The possibility of such a person would pose a threat to the main artery of Raz's argument because the recognition of the independent value of people – and hence, according to Raz, morality – on the one hand, and the prudentially good life on the other, could come apart so that the amoralist might win the 'challenge' against morality. For his life might then be just as good as that of the moral person, and hence he could not be argued into some position by virtue of which he would be committed to the value of people.

So the question becomes: could a person benefit prudentially from a friendship without actually being committed to valuing the friend? Raz thinks not. His argument takes the form of a discussion between him and the amoralist in the following manner: the amoralist starts off from some position which Raz then shows to be mistaken. On the basis of this critique the amoralist then adopts a retreat-

[24] I note that this is consistent with Raz's view on the relationship between values and reasons. As we saw his Classical Conception bases reasons in the value of things. However, according to Raz there are generally three (or four) correct ways to conduct oneself when confronted with the value of something (2001: 161-4): there is reason to Φ if Φ-ing is (1) regarding an object in a way which is consistent with its value, (2) the preservation and non-destruction of, and (3) the engagement with something that has value. Thus valuing friendship does not necessarily commit one to pursuing it.

position which again is undermined, and so on. This argument has six steps, with each step corresponding to a position the amoralist may come to hold and a subsequent critique thereof by Raz. I shall now explain and comment on each of these steps in turn.

For a start the amoralist might adopt the position that the reason why he treats his friends decently is that he wants to do so. However Raz does not think that wanting to do something can be a reason for acting. He presents this claim without argument, instead referring to another article of his, where he does argue that wanting something is not generally a reason for acting.[25] According to Raz's Classical Conception of agency and normativity "[v]alues 'control' reasons in that one can have reasons for an action only if its performance is, or is likely to produce, or contribute to producing, good or if it is likely to contribute towards averting something bad." (p. 47; cf. p. 23) But since there is no necessary connection between wanting something and what one wants being good in any way, a want cannot in and of itself be a reason for action. I do not intend to dispute that wants or desires *qua* mere givens that we are landed with are not reasons for acting, and that we need reasons for satisfying them. But surely the fact that one enjoys doing something can be a reason for satisfying a desire. If so, then if one enjoys treating one's friend decently, one has an agent-relative reason to do so.

Nevertheless, Raz does not consider *this* possible amoralist position. Instead he makes the amoralist proceed by holding the position that he only values his friend because of the special relationship he has with him – namely that he is *his* friend – and not because the friend has any intrinsic value. Raz, however, does not believe that this is a consistent position, for as he argues friendship requires "general concern for the friend as a person" (p. 287), i.e. concern for the friend independently of the

[25] Namely "Incommensurability and Agency", ch. 3 in Raz (1999).

friend's ability to act towards one as one's friend. He explains this general concern with the following example:

"if I and my friend Jane spend our time together, discussing philosophy, going for walks in the hills, and confiding our marital difficulties to each other, then to continue to be Jane's friend I must not only be concerned with her willingness and ability to carry on with the activities which have come to give our friendship its special character. I must also be concerned with her well-being generally." (p. 287)

Raz presents this as a conceptual fact about friendship; to be someone's friend simply is to be concerned about that person's well-being generally, and not just about those features of the friend's well-being which are responsible for the friend's ability to act in the way one appreciates. Furthermore Raz thinks that

"general concern with the well-being of one's friends means that one is treating them as people who have value in themselves, and not merely as people who are valuable to one in one's own life." (p. 287)

Raz seems to believe that these two points – that friendship requires general concern and that general concern in turn entails valuing the friend in himself – should suffice to show that the amoralist is committed to the claim that the friend has value in himself, or agent-neutral value. However, it is clearly not incompatible with agent-relativity that the good of the friend is itself a reason for acting. But this is exactly what Raz seems to presuppose when he claims that being someone's friend commits one to having general concern for the friend – which is true – and that this in turn commits one to his agent-neutral value.

To me it seems entirely consistent to say that I am concerned with my friend for his own sake *because* he is my friend, such that the concern is conditional on the fact that he is my friend. Some people, Raz included, seem to think that friendship is

incompatible with the invocation of this conditional, such that believing in the value of the friendship goes hand in hand with a belief in the agent-neutral value of the friend. I do not see why one's friend cannot be merely agent-relatively valuable (valuable to oneself), while she still is not merely instrumentally valuable, that is, valuable just because it is enjoyable to be around her, or because the relationship enriches one's life. It seems to me completely consistent to say that "my friend has final value *to me*", thereby at the same time believing in her agent-relative and non-instrumental value. Hence, I might believe in her inherent, but not in her agent-neutral value.

Let me explain this point a little further. It seems to me that three distinctions can help clarify this point: the agent-relative/agent-neutral distinction, the intrinsic/extrinsic distinction, and the final/instrumental distinction.[26] The distinction between intrinsic and extrinsic goods has to do with the *source* of the entity's goodness – is the entity good because of some intrinsic features of it, or rather because of some of its external features, that is, its relationship with something else? Hence calling the friendship good is ambiguous between thinking that there are some intrinsic features of the friendship (such as the fact that one enjoys the friend's company) which makes it good, and thinking that there are some extrinsic features of the friendship which make it good (that it improves one's career opportunities, for example).

Likewise there is an ambiguity stemming from the distinction between final goods and instrumental goods. This distinction has to do with the *reason* why the thing is valued – is it valued for its own sake, or for the sake of something else? Thus the friend may be valued for his own sake, or he may be valued only instrumentally, e.g. because he enables one to pursue one's interest in philosophy since he is a good interlocutor when it comes to philosophy.

[26] For the last two distinctions see Korsgaard (1983).

These two distinctions – between intrinsic and extrinsic goodness on the one hand and final and instrumental on the other – together with the distinction already introduced between agent-relative and agent-neutral value allow us to clarify Raz's argument against the amoralist. The point that is crucial is that generally, something can have final value without having agent-neutral value. For example I may clearly enjoy the act of cooking some dish for its own sake – and hence not merely instrumentally – without being committed to the belief that cooking is agent-neutrally valuable.

How does this relate to Raz's claim that general concern for the friend commits one to valuing him in himself? Raz seems to be saying that by virtue of having a friend one is committed to his final and intrinsic value, since first, he is intended to be valued for his own sake, and second – as the example of the friendship with Jane makes clear – the friend is though to be valued because of certain essential properties – in this case a willingness and ability to discuss philosophy and marital problems, as well as being good company whilst hill walking.

Now how about the third distinction, that between agent-relative and agent-neutral value? At this point nothing has been said which should persuade Raz's readers of the agent-neutral value of the friend. That I must treat my friend with general concern in order to be his friend – which in turn means that I have to treat him as someone who has value in himself independently of the friendship, does not commit me to believing that everyone else has a similar reason to treat him in this way, as it would have if I were committed to his agent-neutral value. I might merely believe that the friend has value relative to my life alone. That I place agent-relative value on the friend – that I think he is valuable *to me* – surely suffices to explain why I must value the friend independently of his ability to act towards me as my friend. So on this construal, Raz's argument merely establishes that friendship commits one to the final, intrinsic, and *agent-relative* value of the friend. The distinction between final

and agent-neutral value reveals that one can value the friend in himself, without thinking he has value in himself: one may value him in himself merely *because he is my friend*. If so, then Raz's argument that general concern commits one to the agent-neutral value of the friend is erroneous.

I take this interpretation of Raz's point against the second amoralist position to be entirely compatible with the above quoted passage that "general concern with the well-being of one's friends means that one is treating them as people who have value in themselves, and not merely as people who are valuable to one in one's own life." (p. 287) For although one believes that the friend has value independently of the value one places on him, this does not necessarily entail that one believes that everyone else has reason to value him as well. That I believe that my friend is a good friend, i.e. that he has certain features which make having him as a friend in some way rewarding or beneficial to me, is not only a matter of what he is like but also a matter of what I am like. So if I were to (and could) state the reason why I valued him, my consideration might contain token-reflexives referring to me and certain properties of me. Although rationality requires me to universalize these considerations such that if I believe that something is good for me, it must also be good for any similar person in similar circumstances, this cannot commit me to the proposition that she is agent-neutrally valuable.

First of all, the consideration I use to express my reason may well contain token-reflexives which make reference to me such that universalization becomes impossible. For instance, there may be no one else with the specific historic relationship to my friend which I take to be part of the reason why I value her.

Secondly, the mere fact that the consideration contains token-reflexives means that the value could not possibly be a agent-neutral value, since an agent-neutral value per definition is something everyone has reason to value, no matter how they are constituted, and hence something that does not contain token-reflexives in the

statement of the reason why it is valued. But the amoralist is not necessarily believing that everyone has reason to value the friend – he may merely believe that he himself or (anyone relevantly similar) has. Accordingly, Raz does not manage to show more than that the amoralist is committed to the final and intrinsic value of the friend.

Notice that this is compatible with him believing that the friend is valuable independently of the value he has in his own life. Still, he is not thereby claiming that everyone else ought to value him as well, since his holding that the friend is intrinsically and finally valuable independently, may be something unique to him – an idiosyncrasy, a result of his particular taste people-wise, with the value being dependent upon his preference. If so, then he is not, by valuing the friend, committed to his agent-neutral value, but merely to his agent-relative value – his value relative to him.

Let us proceed therefore on the assumption that Raz has persuaded the amoralist of the final, intrinsic, and agent-relative value of the friend. This, however, does not suffice to make the amoralist committed to the friend being valuable in himself because of his personhood. This is the basis of the third position that the amoralist may come to adopt: now he holds that his friends do have value in themselves, not, however, because they are people, but because they are his friends.

This third amoralist position affirms my interpretation of Raz's statement of the second position as being that the amoralist is committed to the final, intrinsic, and agent-relative value of the friend. For this third position consists precisely in the amoralist pointing out to Raz, that he is not committed to the final value of the friend *qua* person, but merely to the value of the friend *qua* friend. Had he been committed to the value of the friend in virtue of his being a person, then perhaps Raz would have been able to muster an argument to the effect that by virtue of the fact that one

has a friend one is committed to the value of everything which fulfils the definition of personhood, plausibly a term with a broader extension than 'friend'.

Raz attempts to show that the amoralist is committed to the final value of the friend conceived as a person and not merely as a friend, by arguing, firstly, that 'full friendship', which is the kind of friendship in which one cares for the friend for his own sake, requires caring for the friend regardless of whether doing so benefits oneself. This would show that one cares for the friend – insofar as he is a *genuine* friend – as a person, or at least that one cares for him because of some features that are independent of the friendship-relationship.

Secondly, Raz argues that although there may be other forms of friendship – namely 'limited friendships' – in which caring for the friend for his own sake is absent, full friendship is "a good of great importance in human life". This is because, Raz explains,

> "[c]aring for and respecting others […] is important to people's well-being, and it is doubly rewarding and valuable when it happens within a reciprocated relationship, that is, within a (full) friendship." (p. 289)

This second point is the crucial one, since the basic tenet of Raz's argument requires him to show that those who want their own lives to be good – who want to have as high a level of well-being as possible – have a prudential reason to have genuine friendships, and that if they do, they would be committed to the value of people as such. So Raz wants to argue that the amoralist's desire for a good life gives him reason to endorse the kind of friendship in which one values the other *qua* person, and not merely *qua* friend. This requires an argument to the effect that full friendship is qualitatively better than limited friendship. Unfortunately Raz does not supply us with any such argument.

One way of arguing for this will become apparent in the next chapter where I consider Brink's argument that friendship (as well as other relationships with intimates) is a way of extending our own interests. Because we have an interest in developing our deliberative capacities – in learning how to exercise our practical reason well – as well as to extend our interests into the future, we have, Brink argues, a reason to develop what Raz would call full friendships, since they are a means hereto. Since I will give Brink's arguments due consideration later, I shall accept Raz's premise that full friendship is qualitatively better without further examination.

But once we have accepted this conclusion, there is yet a further question, namely whether this or that particular individual has reason to pursue full friendship. Perhaps someone can have enough limited friendships to outweigh several full friendships and on account of a restricted amount of resources, such as time and energy, he might not have reason to pursue full friendships. Another might be too preoccupied with scientific work for the benefit of humanity, leaving him simply unable to fit in such relationships. In cases such as these ones it seems most reasonable to say that *he* does not even have a *pro tanto* reason to pursue full friendship. Such friendships may still be desirable, however, in the sense that *had* he not been too preoccupied with other (worthwhile) activities or already engaged in all those limited friendships, he *would* have reason to pursue them. Therefore it seems that the most reasonable view to hold would be one according to which full friendship, although desirable in and of itself, might not be desirable for this particular individual, vis-à-vis other desirable activities, engagements, and goals, which have already become a meaningful part of his life.

I have already mentioned this crucial distinction between what is objectively desirable and what is objectively desirable for someone in connection with Raz's assumption that there is reason for everyone to desire friendship (as such). Just as it seems false to claim that there is reason for everyone to desire friendship, so it seems

unwarranted to assume that there is reason for everyone to desire full friendships. However, as I pointed out, there may still be reason for everyone to value (full) friendship in Raz's sense, which would give everyone a reason to value whatever a commitment to the value of (full) friendship might commit one to. Let us assume the correctness of this premise to see where Raz's argument will take us.[27]

Raz proceeds on the assumption that the amoralist is committed to the value of his friends *qua* persons, because he thinks that this is entailed by a commitment to the value of full friendship, since here they are valued not only because of their ability to benefit *him*. This in turn makes Raz's imaginary amoralist endorse a fourth position consisting of the claim that it is only because of his *attachment* to the friend that he treats him as a person. In other words, the amoralist's claim is that he only regards the friend as having final and intrinsic value because of his own special relationship to him.

> "The amoralist may while accepting the value of full friendship insist that his friends should be treated by him as people with value in themselves only because they are his friends. He should value them in themselves, and not just in friendship-specific ways. But other people may have no reason to value them at all." (p. 289)

Raz explains the amoralist's fourth position by way of an analogy of an apple tree. We might imagine, Raz proceeds, somebody who is very fond of an apple tree which grows outside his window, because of the fruit it provides, its shade, and the improvement it makes on the view from his window. The fact that this person values the tree, does not commit him to valuing all other trees: "His tree has value for him. Some trees may have value for other people. Some may have value to no one." (p. 289) Likewise with the amoralist's friend: his friend has value for him, some others

[27] Cf. footnote 24.

may be friends of others and be valued by them, and still others may be nobody's friend, and hence be valued by no one.

Raz's reply to this fourth amoralist position has to bring us from the agent-relative value which the amoralist recognizes that the friend has by virtue of his attachment to him, to the agent-neutral value of the friend. Raz believes that there are certain value-presuppositions which commit one to the agent-neutral value of the friend merely by virtue of the attachment:

> "I agree that being attached to an object, person, activity, or project can endow them with a value which they could not otherwise have. They will have that value for people who are attached to them and not for others. But one can only be attached to something if one believes it to be valuable, and the attachment endows the object with extra value only if that object is indeed valuable." (p. 290)

There seem to be two main ideas expressed in this quotation: first of all, being attached to something confers some kind of surplus value on that thing. This surplus value is a special feature of the one who is attached to that thing – it is an extra amount of agent-relative value that arises from the particular way the thing is related to someone because of the attachment. But – and this is the second thought – I could not have endowed the thing in question with additional value in the first place unless the thing was valuable to begin with.[28]

Presumably, 'valuable' here means agent-neutrally valuable. However, this need not be the case since the object being 'indeed valuable', may simply be a matter of the agent not making any reasoning mistakes when he forms the judgement that he finds the thing valuable. In other words it might simply mean that when the agent believes that the thing is agent-relatively valuable, he is not mistaken in his judgement to the effect that this object indeed makes a contribution to his well-being.

[28] Raz explains these propositions further in "Attachment and Uniqueness" in Raz (2001).

That we should indeed settle for the former interpretation is revealed by the fact that Raz, later, after having presented his argument for the claim that one's attachment to something can confer on it extra (agent-relative) value, concludes that

"whatever is intrinsically good for a person is so only if it is good in a non-relativized way, only if it is valuable *tout court*." (p. 291)

Hence, Raz's thought appears to be that in order for one's attachment to something to endow on that thing additional value, the thing must be agent-neutrally valuable.[29] Consequently, if it is indeed the case that attachments are only capable of conferring extra value on agent-neutrally valuable things, and the amoralist grants that this is happening in the case of his friendship – that the friend being his friend makes the friend more valuable in one sense or another – then the value-presuppositions of friendship commits him to the agent-neutral value of the friend. In order to assess this claim properly we need to have a look at Raz's argument for the claim that all attachments – insofar as they confer extra value on the object – presuppose the agent-neutral value of the object. Although it seems to me that object of attachment's being agent-relatively valuable is enough for it to acquire extra value (by virtue of its importance or the special role it comes to play in an individual's life), I will postpone my discussion of this claim until the next section. For now, however, I will for the sake of the argument, grant Raz the conclusion that the amoralist is committed to the agent-neutral value of the friend by virtue of being attached to him.

Upon having recognized that he, by virtue of the attachment, is committed to the agent-neutral value of the friend, the amoralist now adopts the position that since

[29] Moreover, in the aforementioned article Raz says: "The personal meaning of objects, causes, and pursuits depend on their impersonal value, and is conditional on it. But things of value have to be appropriated by us to endow or lives with meaning […]. Attachments are the name I give […] to these appropriations" (2001: 20).

there is no difference between the friend and a tree with respect to value-presuppositions (both are endowed with surplus value because of the attachment), he is still amoral:

"For as long as the argument does not distinguish people from trees there is nothing for him to worry about. His amorality remains intact." (p. 293)

Taking the previous amoralist position into account Raz's idea must be that if attachment to something can convey extra agent-relative value on it insofar as it is agent-neutrally valuable already, the same must apply to the tree. What is difficult to see, however, is how this is supposed to be an objection which the amoralist need consider in the first place. For, after all, the amoralist merely believes in the agent-neutral value of people. So why should he be bothered whether the amoralist believes in the agent-neutral value of trees, or for that matter other material objects as well?

Not knowing exactly what to make of this fifth step, I proceed to the sixth, which seems entirely separate from the fifth. Raz again proceeds from the conclusion reached in step four, that the amoralist's friends cannot be of value merely because he is attached to them. The amoralist can now, according to Raz, endorse the position that

"his friends are people who are valuable in themselves because they are funny, or loving, or whatever other property he values in them. He is still an amoralist because he denies that all people are of value." (p. 295)

The claim is that since it is only because of certain characteristics of the friend that he is valued, such as his wit, the amoralist is not committed to the value of his personhood, but rather him as a person *because* he possesses those properties.

To Raz, however, this amoralist concession opens up a loophole:

"First, the amoralist has, of course, to acknowledge that all people who possess the qualities which his friends have, and which make his friendship with them reasonable, are also of value in themselves. He has, however, to go one step further. He has to admit that all the people with whom it would be reasonable for him to have a personal relationship are also of value in themselves. This is the case even though at present he has no desire to become their friend […]

Second, by parity of reasoning the same goes for those with whom it is reasonable for others to have a personal relationship. If the amoralist believes it is reasonable that Rachel and Robert are friends then he must concede that both of them are of value in themselves." (p. 298)

Accordingly, if the amoralist values the friend because he possesses certain qualities, he must also value any one else who possesses those qualities, and hence grant that those people have value in themselves. Furthermore, Raz maintains, if the amoralist thinks it would be reasonable for him to become some particular individual's friend, then he must acknowledge that *he* also has value in himself, supposedly because the judgement that it would be reasonable to be this person's friend involves a commitment to the claim that it is worthwhile to become attached to this person, which in its turn – given Raz's theory of attachments – involves a commitment to the agent-neutral value of this individual.

Likewise, Raz claims, the judgement that it would be reasonable for someone else to be a third person's friend involves such a commitment because, again, one judges that this person is worthy of the other's attachment. If I believe that someone is worthy of somebody else's attachment, then simultaneously, I believe in the agent-neutral value of that person. At least that is what Raz thinks follows from his claims about attachment.

Raz concludes that the amoralist is committed to something that he calls *the potentiality principle*, viz.

"it is possible for all people to possess the qualities which make people valuable in themselves (being witty, wise, good-looking, etc.), when that is understood to mean that they belong to a species of animal which can, consistently with their nature as members of that species, possess these qualities." (p. 300)

So other people could have been valuable in themselves, since they could, counterfactually, have possessed the qualities which make people valuable in themselves.

Raz ends his paper with posing some doubts as to whether the moralist has anything stronger than the potentiality principle in mind. I will not enter into these considerations since I have already said what I take to be involved in accepting morality. So let me simply summarize what I take to be the six steps of Raz's argument:

(1) The amoralist cannot treat his friend as a friend merely because he wants to treat him in that way, since wanting to act in a certain way "is not the sort of thing that can be a reason." (p. 285)
(2) Since friendship requires general concern for the friend, i.e. concern for his well-being independently of his ability to act towards one as one's friend, the amoralist cannot merely value his friend because of his special relationship to him. The amoralist must grant that the friend has value in himself.
(3) The amoralist cannot hold that the friend has value in himself merely *qua* friend and not *qua* person. For this would amount to a limited friendship which makes less of a contribution to the well-being of the amoralist than full friendship does.
(4) The amoralist cannot consistently hold that the friend has value in himself merely because of his being attached to him. For in order for attachments to have value –

in order for them to make a contribution to the well-being of the amoralist – the object of the attachment must be agent-neutrally valuable.

(5) The amoralist cannot consistently hold that the object of his attachment in the case of material objects has agent-neutral value just like in the case of persons, since relationships with persons are reciprocal.

(6) The amoralist can hold that his friend has agent-neutral value because of certain properties which make him worth being attached to, but which may not be shared by others. However, it is not clear that the moralist has anything stronger in mind.

Since I am particularly interested in the move from the agent-relative value of things to their agent-neutral value, in the following I shall focus on the fourth step.

2.3 THE COMMITMENT TO THE AGENT-NEUTRAL VALUE OF PEOPLE

In the previous section I summarized and explained the six steps I take to make up Raz's argument in "The Amoralist". Of these steps (3) and (4) seems to be the most contentious. If (3) is to hold, then an argument to the effect that full friendship is qualitatively superior to limited friendship has to be produced, such that it, from the point of view of prudence, always would be more desirable than limited friendship. Someone might, as we shall see Brink in fact does in the next chapter, give good reasons as to why this is so, for it might well be that one develops one's deliberative capacities best within the context of a full friendship.

Since my main concern is the transition from agent-relative to agent-neutral values signalled by step (4) I shall not pursue this question in any further detail at this point. The fourth step, as we saw, involves the amoralist claiming that it is only because of his attachment to the friend that he is considered to be (agent-relatively) valuable. Raz's reply to this was that the agent-relative value of the friend must be viewed, as

it were, merely as an extra, idiosyncratic portion of value on top of the agent-neutral value which must be presupposed if the agent's attachment to the friend is to have obtained this surplus value in the first place. Raz's strategy, as we saw, was to propose an intimate connection between what is agent-relatively valuable and what is agent-neutrally valuable, such that

"whatever is intrinsically good for a person is so only if it is good in a non-relativized way, only if it is valuable *tout court*." (p. 291)

Because Raz takes this to be a conceptual fact, the claim must be that the agent cannot consistently believe that his attachment to his friend is agent-relatively valuable without also believing that it is agent-neutrally valuable. For understanding what it is for one's attachment to somebody to be in a special way valuable to oneself, means understanding that the object of one's attachment is agent-neutrally valuable.

Raz presents this claim as stemming from a conjunction of three principles: firstly what he calls the "reciprocal relationship between good (or valuable) and good (or valuable) for one." (p. 290) Secondly, the thesis that "all intentional action is undertaken for a reason" (p. 291), and thirdly the reality principle, which means that "our intentional actions (including those that manifest our attachments) are undertaken for what we believe to be good reasons, they are worthwhile – unless accidentally – only if those reasons are sound." (ibid)

The first of these principles seems to be the most controversial, and I would like to try to clarify further what is meant by it. I assume Raz has in mind the distinction between agent-relative and agent-neutral value, where agent-neutral value corresponds to 'good', 'good in a non-relativized way' or 'valuable *tout court*' and agent-relative to 'good for one' in Raz's vocabulary.

Here is how Raz explains the thesis of the reciprocal relationship between good and good for one:

"On the one hand, nothing can be good unless it is possible for it to be good for someone. We can imagine goods which are not actually good for anyone at the moment, or even goods which are unlikely to be good for anyone in the foreseeable future. But it is unintelligible to say of something that it is good or valuable if it is impossible that it be of value to anyone. On the other hand, if anything is good and one can relate to it in 'the appropriate way' then other things being equal it is good for one. So if a novel is a good novel and I can read it with understanding then it is good for me to read it, other things being equal. There is no more to something being good for me than that it is a good which is within my reach (though of course other things may be better, or the cost of reaching it may be more than it is worth, etc.)" (p. 290-1)

I shall not embark on a detailed discussion of the first part of the reciprocity thesis, which is irrelevant to Raz's claim that something being valuable *to an agent*, commits that agent to its agent-neutral value. For the first part merely states that 'nothing can be good unless it is possible for it to be good for someone', which seems to be equivalent to 'something cannot be agent-neutrally valuable unless that thing has the potential to be an agent-relative value' or what is the same 'if it is impossible for X to be agent-relatively valuable, then X cannot be agent-neutrally valuable', which in turn is the contra-positive of (I).

(I) 'if X is agent-neutrally valuable, then it is possible for X to be agent-relatively valuable'.

The second conjunct of the reciprocity thesis is somewhat harder to pin down from the above quote in which Raz supposedly puts it forth. The problem is that the second part of the quote starting with 'On the other hand' seems to contain two distinct theses. First, we find 'if anything is good and one can relate to it in 'the appropriate way' then other things being equal it is good for one', which seems to have the same meaning as

(II) 'if X is agent-neutrally good and Y can relate to X 'in the appropriate way', then, other things being equal, X is agent-relatively good for Y'.

Secondly we find: '[t]here is no more to something being good for me than that it is a good which is within my reach (though of course other things may be better, or the cost of reaching it may be more than it is worth, etc.)' means that X's being agent-relatively good for Y is a sufficient condition for X's being an agent-neutral good which is within Y's reach, by which we get

(III) 'if X is agent-relatively valuable to Y, then X is an agent-neutral good that is within Y's reach'.

Although (I), (II), and (III) are all implied by what Raz refers to as the reciprocity thesis, I will focus my discussion of it on (III), since that is the one that is relevant to the present investigation. (III) would, if correct, allow us to infer that the amoralist is committed to the agent-neutral value of his friends, since otherwise, according to this principle, the friend would not have been agent-relatively valuable in the first place. This is because being agent-relatively valuable is all that is required for something's being agent-neutrally valuable. But Raz does not argue for this crucial claim of his. He does, however, make a reference to another paper of his, 'On the Moral Point of View', where there is some mention of the reciprocity thesis. Here it seems that something by way of argument for the thesis is present:

"Typical explanations of what is good for agents display a dual structure. It is good for Johnny to play the piano because engaging with music is good, and because he has enough of an ear and physical control to be able to do so by playing the piano. In other words in justifying that anything is good for any agent we show (a) that the thing is good, and (b) that the agent has the

ability and the opportunity to have that good. I can think of no other way to account for why what is intrinsically good for some person or other […] is good for them. If this is so then relational goods presuppose non-relational goods." (p. 260)

However, the fact that we typically cite the agent-neutral value of some activity or object when we *explain* or try to comprehend why something is agent-relatively valuable for somebody does not show that he is *committed to* the agent-neutral value of the activity or object, when he values it. For firstly the conclusion does not follow from the premise: the fact that we often refer to the agent-neutral value of things in order to grasp why something is good to people, does not establish that it is a conceptual fact about the person that *he* by virtue of valuing e.g. music is committed to the agent-neutral value of music. But a second and related point is that not all agent-relative values can be explained according to Raz's schema in the first place. Examples such as counting the blades of grass on various lawns illustrate this. Here, supposedly, we cannot explain why it is a value *to them* by making reference to something such as the usefulness of knowing the exact amount of blades of grass on this and that lawn to everyone else. Although we do at times explain why something is good to an agent by saying that it the thing is good and that the agent has the ability and the opportunity to have that good, it also appears to be a completely legitimate explanation to cite merely the fact that the thing is good *to him*, i.e. that he has the ability to appreciate that good. Thus if someone enjoys counting blades of grass, and if it brings him satisfaction then it seems we will have explained the value *to him* of this activity without making reference to its agent-neutral value – after all, it might not be agent-neutrally valuable at all (Raz would probably agree with this).

But, and this is my first point resurfacing, even if *we* as observers needed to cite agent-neutral values in order to make sense of why someone valued something, it would appear to be a second matter to show that the agent is committed to the agent-neutral value of what he is valuing, as he might not value it for that reason. If I value

playing music, I am not committed to believing that everyone else has reason to value it is well – I just like it, that is all. It seems therefore that it is entirely satisfactory to justify that it is good for me to play music to cite merely the fact that it is good *for me*, i.e. that I enjoy it.

It seems therefore that (i) we do not need to *cite* agent-neutral reasons in order to explain why something is good for an agent, (ii) nor do we need them to justify something being good for them, i.e. agent-relatively valuable. Since therefore Raz fails to argue for claim (III) of the reciprocity thesis, there is no justification for believing in a necessary connection between someone valuing something and him being committed to its agent-neutral value. Accordingly, Raz fails to show that the amoralist is committed to the agent-neutral value of the friend.

As will be remembered, Raz presented the argument for step (4) as consisting of three principles, and so far only one – the reciprocity thesis – has been discussed. Perhaps Raz's remaining two principles can alleviate the defectiveness of the first principle? The second, as will be remembered, said that all intentional action is action undertaken for a reason, which Raz takes to be a "belief that it [i.e. the action] or some of its consequences are good." (p. 291) But obviously 'good' here is ambiguous between agent-relative and agent-neutrally good, and Raz does not argue that one must believe in the agent-neutral goodness of the action or its consequences when one acts.

Accordingly, this principle cannot as it stands be used in support of the view that one implicitly places agent-neutral value on the friend when one acts for the sake of him. That would have been the case if Raz's second principle should be interpreted as 'all intentional action is action undertaken for a reason, i.e. a belief in the agent-neutral value of either the action itself or its consequences', for in that case one could argue, via (II), that since the amoralist is appropriately related to the friend, he would, merely by virtue of standing in a relation of friendship to someone else, be

committed to the agent-neutral value of him. For since (full) friendship involves acting for the sake of the friend's good, this would have enabled Raz to conclude, granted our current interpretation of the second principle, that the amoralist is committed to a belief in the agent-neutral value of the friend.

However, besides being quite implausible, Raz does not lend any support to this interpretation of this second principle. Perhaps he deliberately intends the second principle to be ambiguous in this way between agent-relative and agent-neutrally valuable, since he believes the first principle to be correct and therefore able to warrant the desired conclusion in conjunction with the two other principles.

The third and final principle – the reality principle – means that,

> "as our intentional actions (including those which manifest our attachments) are undertaken for what we believe to be good reasons, they are worthwhile – unless accidentally – only if those reasons are sound. That implies that only if the objects of our attachments are of value is there any value in our attachments, or in the actions which manifest them." (p. 291)

Raz's thought here clearly seems to be that in order for an attachment to be valuable to an agent, the object of the attachment must be agent-neutrally valuable – valuable independently of the attachment; only if the reason why we form the attachment is a good reason – only if the reason is in fact a reason for forming the attachment – is there any value in the attachment. However, this clause seems obviously false. Someone may form an attachment to a completely valueless person, without the attachment itself being valueless to her. Imagine, for instance, a woman in a relationship to a guy – a complete rat who mistreats her. Despite the mistreatment she may find that the relationship gives her life some kind of meaning and content, sad though it is. Further, she might well realize that he is totally without merit. But even so, that does not make the relationship without value to her – it might have taken on a special significance to her, and she might even say that her life would be

empty without him. Hence, it does seem that there can be value in the attachment without the object of the attachment being valuable.

But besides this, Raz's argument seems to be a *non sequitur*. For even if we grant that we undertake our actions for what we take to be good reasons, and that our actions and attachments are worthwhile only if those reasons indeed do support the actions and attachments in question, it would not follow that the *objects* of our attachments have to be valuable in order for the attachment to have value. The conclusion which in fact follows from the premises is that the attachment itself – and not the object of the attachment – has to be valuable if the attachment is to have value. Hence, Raz's third principle is unable to show that the amoralist is committed to the value of the friend.

I therefore conclude that none of Raz's three principles can be used to support (4). The first as we saw relied on the claim that (III) 'if X is agent-relatively valuable for Y, then X is an agent-neutral good that is within Y's reach'. This claim however seems to be gainsaid by common examples of engagements with activities, or attachments with objects, that are without any agent-neutral value. In addition, Raz's argument for this clause seemed to be operating on the level of explanation and not on the conceptual level. In other words the claim seemed to be that when we try to comprehend the meaningfulness of others' attachments and pursuits we often cite the fact that the object or activity is in fact valuable and worthwhile. But what is really needed if Raz's argument was to work was the conceptual claim that valuing something commits one to the agent-neutral value of that thing, in the same way that sincerely believing that something is red commits one to the belief that it has a colour. The second and third principles likewise seem unable to support (4). The second because it traded on a dual meaning of 'good' when it claimed that all intentional action is undertaken for a reason, which to Raz is a belief that it or some of its consequences are good. And at any rate we do not perform all our actions because

we think there is something particularly agent-neutrally good about them. The third principle, that our attachments must, to be valuable, be attachments to something that is valuable, seems to be wrong. Some attachments do carry a special significance and importance in people's lives although they are attachments to utterly valueless people or objects.

2.4 CONCLUSION

In this chapter I have assessed Raz's attempt to argue the amoralist into accepting something akin to morality by investigating the value theoretical presuppositions behind friendship. In section 2.1 I explained that Raz's amoralist is best thought of as someone endorsing the principle of the rational egoist, i.e. that if some action constitutes or brings about a benefit to some agent, this gives him a reason for perform the action. I then continued to explain and assess Raz's argument. In section 2.2 I summarized Raz's argument as consisting of six distinct steps, through which Raz attempted to persuade the amoralist of the agent-neutral value of his friend, which in turn would commit the amoralist to something akin to morality.

However, it seemed entirely consistent for the amoralist to merely believe in the final and agent-relative value of the friend. Raz argued, however, that the agent-relative value placed on certain entities such as friends, brings with it a commitment to the agent-neutral value of these entities.

In section 2.3 I assessed this claim and the argument purporting to establish it and found it unpersuasive. Raz's argument for the claim that attachments, in order to have agent-relative value, must have agent-neutral value, and that the amoralist is committed to this in placing agent-relative value on the friend, was presented by Raz as relying on three claims. These three claims were considered in turn and found wanting. The first of these – the so-called reciprocity thesis – was found to be mistaken, since the argument purporting to establish it was misguided. It seemed to

rely on the mistaken idea that one has to cite the agent-neutral value of something in order to account for its agent-relative value. The two other claims – the claim that intentional action is action for reasons, and the reality principle – were also shown to be flawed.

Therefore Raz does not succeed in establishing that the amoralist is committed to the agent-neutral value of the friend, and unsurprisingly therefore, his attempt to argue the amoralist into morality fails. It might be correct – consistently with Raz's Classical Conception – that reasons are based in the value of things, and further that the well-being of a person should be understood in terms of whether it has certain objectively valuable goals, activities, etc. in it. Nonetheless, although friendship might be one such objectively valuable pursuit, one cannot derive agent-neutral reasons from a commitment to its value, since as I have argued one need not believe in the agent-neutral value of the friend to begin with.

3. BRINK'S APPROACH: EUDAIMONISM

In this chapter I want to investigate a contemporary version of eudaimonism, namely that of David Brink. In a paper he sets out to investigate to what extent prudential reasons – that is, reasons arising in our concern for our own well-being – can be shown to recommend the same actions as those that are demanded by morality. He does this by arguing that people can see each other's interests as extending their own, which thereby gives them a *pro tanto* reason to take the interests of others into account, because acting for the good of others is a way of benefiting themselves. Hence, according to Brink, if people reflect properly on the matter, they will come to see that heeding moral demands is in their own self-interest.

In section 3.1 I want to say something about how Brink's eudaimonism, which gives an egoistic justification of morality, relates to other such attempts, in addition to the two previously discussed attempts. Brink's egoistic justification of morality is distinguished from other egoistic accounts, such as that of Hobbes for instance, in that it is based on certain metaphysical assumptions about what it is to be a person. But all kinds of egoistic justifications of morality are fundamentally different from the views we have been considering so far. Egoistic justifications essentially proceed by giving agent-relative reasons as to why one should be moral, and hence try to account for morality's other-regarding aspect in terms hereof. They therefore deny that morality is even partly agent-neutral – that it consists to some degree of considerations about what there is reason for anyone to do irrespective of his particular projects and relationships – something which was upheld by the two previous approaches.

In section 3.2 I state and discuss what I take to be Brink's argument, and in section 3.3 I consider various objections to this argument, one of which I take to be final. The objection states that although Brink manages to justify a large proportion of morality on an eudaimonistic, purely agent-relative basis, he encounters problems in the most

notable impartial value recognized by common sense morality: justice. Since an account based solely on agent-relativity is unable to account for this element in morality, I will argue that Brink's approach is unsuccessful. This lends support to my thesis that any attempt to justify morality must appeal directly to agent-neutrality, and conversely, that one can not be argued into morality from a purely agent-relative starting point.

3.1 METAPHYSICAL EGOISM

Hobbesian contractualism argues that there is reason to be moral by assuming a prudential conception of rationality – roughly that there is reason to perform an action if and only if it results a benefit to the agent – and arguing that this renders moral conduct rational. Hobbes saw morality as a useful convention derived from an envisaged initial state of nature, characterized as a state of war "of every man, against every man." (L, XIII, 8)[30] This war of all against all Hobbes thought to be upheld by three factors: competition, "diffidence" (or fear) and want of "glory" (or reputation) (L, XIII, 3-9). Competition, which is for the basic necessities of life, makes us struggle with other human beings; fear of others will make us fight them so as to ensure our own safety; getting a reputation as someone who is not violated against with impunity, will have protective benefits. But – it goes without saying – acquiring such a reputation takes effort. These three factors, together with the – all things considered – equality of man, turns the state of nature into a war of all against all.

According to Hobbes, the state of nature is something that ought to be avoided. Because people have a desire to avoid death and for a "commodious living" (L, XIII, 14) they ought to leave it behind. Hobbes thus seems to base his argument for the reasonableness of morality – which is what is chosen as the alternative to the state of

[30] References are to chapter and paragraph number in Richard E. Flathman and David Johnston's edition of *Leviathan* (New York: W. W. Norton & Company, Inc., 1997).

nature – upon a principle of practical rationality, which has been called "the preservationist account of rationality" (Gert's term in (1989)). This account maintains that it is rational to act on those desires which lead to one's long-term benefit, primarily those that results in the preservation of one's own life. Hobbes assumes that we act from a settled desire to preserve our own lives, and takes morality to be instrumental to this end, which is to say: we ought to be moral since this is for our long term benefit.

The uncertainty and danger of the state of nature renders it rational to leave it behind, and choose to accept certain constraints on one's actions – say, constraints against stealing from and injuring others – granted that others impose the same constraints on their conduct. Thus morality, according to this Hobbesian contractualist view, is fundamentally a mutually advantageous convention. This view in turn yields one possible answer to the normative question which attempts to show the rationality of moral conduct: we have a prudential reason to adhere to the demands of morality because, ultimately, we benefit ourselves by doing so.[31]

Eudaimonists share the belief that moral actions can be justified on a solely prudential conception of rationality. Being eudaimonists they hold that an agent's practical reasoning should be guided by an objectively correct conception of

[31] There are other interesting versions of contractualism, which I unfortunately cannot discuss in any detail here (and I do not mean to suggest by this that my brief discussion of Hobbesian contractualism is fully satisfactory either). Gauthier's version is distinguished by the fact that it starts from the distinctively post-modern belief that we no longer believe in objective moral values (cf. Gauthier, 1991), and it therefore assumes that we cannot appeal to such values when justifying morality. Gauthier also assumes something like the prudential conception of rationality (ibid. 19), and in effect tries to *reduce* morality to rationality – morality *just is* what the rationally justifiable modes of conduct that follow from the assumed conception of rationality. What is known as Rawlsian (or Kantian) contractualism on the other hand, attempts to argue for the rationality of moral conduct without ignoring moral considerations altogether, in that it assumes that people are motivated by a desire to be able to justify publicly their claims on others.

eudaimonia – that is, that the ultimate end any given rational agent has reason to pursue is his or her own happiness (which is the, perhaps unfortunate, stock translation of the Greek word *eudaimonia*). Since eudaimonists furthermore regard possession and exercise of the moral virtues as an essential part of eudaimonia, they are committed to the view that possession and exercise of the moral virtues is an essential part of happiness, or the good, flourishing life.

Consequently, in their attempts to provide an answer to the normative question, eudaimonists proceed in the same way as Hobbes: they attempt to show that adherence to the demands of morality, specifically, that possession and exercise of moral virtue is an essential part of the prudentially good life. The manner of addressing the normative question is basically the same: first one takes the prudential conception of practical rationality to be correct, and then one tries to show, by resorting to this conception, that there is indeed a prudential reason to act according to moral demands or to become virtuous. In other words, what there is agent-relative reason to do, coincides with what there is moral reason to do.

Hobbesian contractualism and eudaimonism thus share the belief that morality is not fundamentally about what there is agent-neutral reason to do. The consideration that makes up the reason to be moral involves an essential backward reference to the agent himself. Morality is about what there is agent-relative reason to do for the agent – if someone values *his own* eudaimonia then there is reason *for him* to acquire the virtues, that is, to act morally from a settled disposition, because this is a way of improving his eudaimonia. Hence the claim is not the Korsgaardian one that there is an agent-neutral reason to value humanity wherever it is found. Nor is the claim that there is reason to be moral because, as Raz thought, persons are agent-neutrally valuable. Rather, there is reason to be moral because it is a way of securing one's own

interest in a flourishing life. We find this attempt exemplified in the works of the two main historical exponents of eudaimonism: Plato and Aristotle.[32]

In this chapter I shall assess one contemporary form of eudaimonism which I have found to deserve special mention: that of David Brink. In his article 'Self-love and Altruism', he proposes an approach to the normative question which draws on the eudaimonism of Plato and Aristotle together with some insights from the philosophy of the British idealist T. H. Green.[33] Brink dubs his theory 'metaphysical egoism', whereby he intends both to distinguish it from what he calls 'strategic egoism' and also to draw attention to the fact that this view is based on certain metaphysical observations about persons.[34] By strategic egoism Brink basically understands the kind of theory that we have already seen exemplified in Hobbes' contractualism, i.e. the type of theory that tries to square prudential and moral reasons by assuming that "different people's interests are conceptually distinct but [which] argues that they are in fact causally interdependent." (p. 123)

However, according to Brink, there is little reason to believe in the prospects of this kind of approach. He states three objections to it. Because strategic egoism justifies moral obligations towards others in terms of the different parties' bargaining power, it cannot account for certain obligations considered a part of common-sense morality. First of all, it cannot justify duties towards future generations who have no possibility of harming us today.

Second, extreme inequalities in natural endowments and resources would mean that those who are at the one, best off extreme would have no reason to bargain with those at the other, worst off extreme. Hence, "if the wealthy and talented have sufficient strength and resources so as to gain nothing by participating with the weak

[32] Here I follow the *standard* way of interpreting Plato and Aristotle. Cf. T.H. Irwin (1995).

[33] Unless I indicate otherwise, parenthetical references will be to this article.

[34] Cf. p. 124.

and handicapped in a system of mutual cooperation and forbearance, the former can have no reason, however modest, to assist the latter." (p. 123)

Thirdly, the rationality of compliance with moral duties is conditional upon the possibility of being detected in the sense that if it is very unlikely, or even impossible that failure to comply will be detected, then there is reason not to perform one's duty, because one could thereby receive the benefits from others' compliance without having to sacrifice one's own self-interest. However as Brink points out "moral norms seem counterfactually stable – they would continue to apply in these counterfactual circumstances – as other-regarding norms that the strategic egoist can justify are not." (p. 123) Brink is certainly right about this; duties are not, according to common-sense morality, overridden by the fact that failure to comply goes undiscovered.

Metaphysical egoism, conceived as a philosophical position, is distinguished from this kind of approach to the normative question by virtue of the fact that it regards the interests of people as "metaphysically, and not just causally, interdependent such that acting on other-regarding moral requirements is a counterfactually reliable way for an agent to promote his own interests." (p. 124) In the following section I turn to an exposition of Brink's main argument for this bold claim.

So far, I have explained how Brink's version of eudaimonism – metaphysical egoism – relates to the view of morality, implicit in the views which we have been considering so far. Brink, just like Hobbes, denies that morality essentially involves claims about what there is agent-neutral reason to do. Instead, the task becomes that of showing that what there is agent-relative reason to do as a matter of fact is the same as what we intuitively think there is moral reason to do. Strategic egoism, as we saw, failed in this task – it was unable to account for the rationality of moral action towards others who were without bargaining power and in cases where one could act immorally without being discovered. In the following section I summarize and

explain Brink's argument, and in the subsequent one, I consider whether his version of egoism does a better job of justifying morality.

3.2 BRINK'S ARGUMENT

Before I said that Brink's position – metaphysical egoism – derives its name from the fact that it is based on certain metaphysical claims about persons. What are these?

According to Brink, it is essential to being a person that one is able to exercise capacities for practical deliberation – that one is able to formulate, assess, revise, choose, and implement "projects and goals in light of a conception of what is best" (p. 136). The next question we will need to ask is what it means for a person to persist or for him to continue to exist in the future. Under what circumstances can we say that a person is the same person? Brink argues that personal identity must consist in some kind of continuity of mental life: "For it is only those physical changes that destroy continuity of mental life that destroy a person; other physical changes are alterations in a persisting person." (p. 136)

In support of the claim that personal identity is primarily a function of mental – as opposed to physical – continuity, Brink brings forth certain thought experiments in which the mental and physical aspects of a the same person's mental life are severed, and in which our intuitions about personal identity seem to track the mental life rather than the physical life. In one such thought experiment – described by Sydney Shoemaker – surgeons are able to remove and transfer the brain from one person over into the body of another, thereby moving his mental life along with it. As Brink points out we would clearly think that the person who has had his brain removed and put into the head of another body *is* now the same person in this (other) body: We would not think that the person previously occupying this second body is still that person now that his brain has been removed from it, because now only his body

is left. Therefore, the persistence of a person seems to supervene on the persistence of mental, as opposed to physical, life.

On the basis of these observations Brink concludes that

"what makes persons at different times the same person and, hence, what unites different parts of a single life is psychological continuity. A series of persons is psychologically *continuous* insofar as contiguous members in a series are psychologically well connected. A pair of persons are psychologically *connected* insofar as the intentional states (e.g. beliefs, desires, and intentions) and actions of one are causally dependent upon those of the other." (p. 138)

By means of a second thought-example, Brink tries to illustrate the fact that it is psychological continuity and not personal identity that is the primary determinant of the rationality of concern. That is, it is not the fact that A is *identical* to B that makes it rational for A to be concerned with B – rather, it is the fact that A is *psychologically continuous* with B that makes it rational for A to be concerned with B. Hence insofar as it can be argued further that one is psychologically continuous with other people, this point would have obvious consequences for the rationality of other-regarding concern.

This second thought-example, intended to show that concern (if rational) is a function of psychological continuity, asks us to imagine three identical triplets – Tom, Zeke, and Zach who are in a serious car accident, leaving Zeke and Zach are both brain-dead, while Tom's brain survives, although the rest of his body is completely destroyed and useless as a result of this accident. Brink now describes three possible scenarios: in case 1 Tom's brain is transplanted into Zeke's body, which preserves his (Tom's) psychological continuity. Here we regard Tom as the surviving recipient and Zeke as the dead donor. In cases 2 and 3 Brink assumes that it is possible for half of Tom's brain to sustain psychological continuity with Tom. In case 2 it is assumed that

only half of Tom's brain survives the accident, and that this half is transplanted into Zeke's body. Again we would, Brink concedes, claim that Tom survives.

In case 3, it is assumed that the whole brain is intact, but here we transplant one half into Zeke's body, and rechristen him Dick – and the other half into Zach's, whom we now call Harry. This is a case of what Brink calls fission; and he assumes, for the sake of the argument that there is the same amount of psychological continuity between Tom and Dick and Tom and Harry, as there was between Tom and the recuperating patient in cases 1 and 2. But if this is so, Brink argues, identity is not a necessary condition for the rationality of concern. This is because – although it is rational for Tom to be concerned about whoever he will turn out to be after the surgery (Zeke in cases 1 and 2, Dick and Harry in case 3) – it would be wrong to suggest that Tom is identical to Dick and Harry in case 3.

For first of all, it is not, according to Brink, plausible to say that Tom survives as one rather than the other. We have no grounds for giving preference to identifying Tom with any one of his "psychological relatives" to the other, for both seem to have an equal claim to being Tom.

Second, it would be false to say that Tom survives as both of them at the same time, because identity is a transitive relation, such that if Tom = Dick, and Tom = Harry, then Dick would have to be identical to Harry. But, as Brink points out, Dick and Harry have completely different properties: they have different streams of consciousness and wake up in different beds, and hence cannot be identical. Therefore neither can Tom, given the transitivity of identity, be identical to both of them.

Neither – which would be a third option – can Tom survive as the scattered person consisting of Dick and Harry. To Brink this suggestion simply does not make sense because "persons must be functionally integrated systems; if one is to be held responsible for one's actions, then one's actions must be caused in the right way by

one's beliefs, desires, deliberations, and choices." (p. 140) This, however, obviously does not apply to Dick and Harry; they are not functionally integrated: for instance, an intention formed in Dick does not cause an action in Harry. Hence it would be wrong to regard Tom as being one person consisting of both Dick and Harry.

By way of this elimination of the possibilities, Brink concludes then that it is psychological continuity and not identity that warrants the rationality of concern. Tom should regard, even in the case interpersonal (as opposed to in*tra*personal in cases 1 and 2), the other people as extending his interests:

> "by virtue of being fully psychologically continuous with Tom, Dick and Harry will each inherit, carry on, and carry out Tom's projects and plans (though presumably in somewhat different directions over time). This seems to be a good ground for claiming that Dick and Harry extend Tom's interests, in the very same way that his own future self would normally extend his interests […] This helps us better understand the common claim, which Plato, Aristotle, and Green all endorse, that in more conventional interpersonal cases there is interpersonal extension of interests. Among intimates they claim, B's good can be regarded as a part or component of A's good. The ground they offer for this claim is that A and B interact and help shape each other's mental life; the experiences, beliefs, desires, ideals, and actions of each depend in significant part upon those of the other. These are the sorts of conditions of psychological continuity and connectedness that are maximally realized in normal intrapersonal cases and in fission cases. Here they are realized to a very large extent in familiar interpersonal cases. This means that each should regard the good of those to whom she stands in such relationships as a constituent part of her overall good." (p. 142-3)

So in other words the mutual shaping of one another's mental life that takes place among intimates (friends, spouses, relatives, parents and their children, etc.), is analogous to a brain transplant in respect of mental life. By interacting with intimates – by shaping the other's mental life – one develops a psychological continuity with this person. As Brink explains, this takes place because of the transformation of

experiences, beliefs, desires, ideals, and actions, which ensues from the discussions, exchange of ideas and points of view that are characteristic of such relationships.

Brink takes this to be a justification for concern for this person, i.e. other-regard expressing concern for this other person's well-being. Insofar as one's intimate is thriving, one is thriving oneself, because him thriving involves having one's own interests considered, since intimates share in each other's interests. Hence one has an egoistic reason to be concerned about the intimate because this is instrumental to being concerned about one's own interests.

But in a certain sense it is still relevant to ask why it is worthwhile for the agent to be concerned with the good of someone else although he extends one's interests. For obviously this may require taking away some resources – financial, emotional, and intellectual – from somewhere else and using them in pursuit of the intimate's well-being. Assuming the agent is motivated by concern for his own eudaimonia first and foremost, why would it be rational for him to expend resources on the development or continuation of an intimate relationship with someone if he could have spent those resources directly on himself? The answer, according to Brink, is that a single person is not, on his own "self-sufficient at producing a complete deliberative good." (p. 144) For as he explains:

"Insofar as we regard the exercise of deliberative capacities as the chief ingredient in *eudaimonia*, we can see how self-understanding and self-criticism are both parts of *eudaimonia*. Interaction between those who are psychologically similar provides a kind of mirror on the self. Insofar as my friend is like me, I can appreciate my own qualities from a different perspective; this promotes my self-understanding […] But interaction with another just like me does not itself contribute to self-criticism. This is why there is deliberative value in interaction with diverse sorts of people many of whom are not mirror images of myself. This suggests another way in which I am not deliberatively self-sufficient. Sharing thought and discussion with others, especially about how to live, improves my own practical deliberations; it enlarges my menu of options, by identifying new options, and helps me better assess the merits of these options, by forcing on my

attention new considerations and arguments about the comparative merits of the options." (p. 145)

It is in my interest to exercise my deliberative capacities, and they are developed in part by engaging in certain forms of relationship in which I can see my own qualities from a different perspective and thereby promote my self-understanding, and come to criticise my own beliefs, behaviour, responses, and other features.

It certainly seems correct that it is easier to understand oneself when one sees one's own qualities represented in another. Perhaps because one can observe in this other individual the consequences of certain character-traits he shares with oneself, which enables one to gain a fuller understanding of what is involved in having that trait. But it is not clear to me that the friend has to be different from me in order for such observations to generate self-criticism. If I, for instance, share the properties of being arrogant and sarcastic with a friend, and I observe the way this is hurting others who suffer the consequences of these characteristics, what else is required for me to be self-critical? All that seems to be needed is that I become aware of these consequences.

However it does seem to be correct that being confronted with someone who is unlike me can contribute to self-criticism. One common way in which this happens is that I observe somebody who acts in a way which I think – taking the factors of the situation into account – is contrary to what there was reason for him to do. Being puzzled by his actions may lead me to realize – perhaps through conversations with him – that he was right; that his sensitivity to the reasons which bore on the situation and/or his ability to balance these reasons against each other was superior to mine. Similarly with inferior (practical) reasoners: they will be unable to come up with considerations which count in favour of what they did, thereby reassuring me that I was right about them being wrong. In this way, one's being able to criticise oneself, and hence to develop one's susceptibility to and knowledge of the considerations

which count in favour of (or against) certain actions, as well as their relative importance, seems to be connected with getting in contact with people who are dissimilar from oneself.

So it seems correct for Brink to claim that relationships – be it with likeminded or disparate individuals – can contribute to improving deliberative capacities. But the crucial question is: Why should it be in my interest to develop these capacities, or alternatively: how can concern for my own eudaimonia, render it rational for me to try to develop my deliberative capacities? Might I not be perfectly content and happy without developing my deliberative capacities to the full of my potential? Of course, the more developed one's deliberative capacities are, the better one may be able to achieve one's goals, but Brink's claim goes beyond this. He claims that developing one's deliberative capacities also helps to develop the goals themselves: it enlarges my menu of options, and helps me to assess the quality of them, as he says. It seems obvious how this would work: Through interacting with other people I may come to realise that some of my pursuits and opinions are shallow and frivolous – upheld only because of prejudice, tradition or the influence of mass culture.

Hence, the exercise of deliberative capacities is not merely instrumental to already existing goals – by doing so one discovers new aims and pursuits, which may be superior in quality to the aims and pursuits one is already engaged with. If there are indeed such qualitatively superior activities and aims, then we can see how the agent has a prudential reason to exercise and develop his deliberative capacities, since that would be a way of opening his eyes to these goals.

Besides this argument, Brink seems to have an argument for the intrinsic value for the agent of the exercise of his deliberative capacities. This argument consists in an invocation of Aristotle's so-called *ergon*-argument in book I, chapter 7 of the Nicomachean Ethics (= *NE*). Here Aristotle defines the chief good for human beings by reference to the human function (*NE*, 1097b22-33), which is taken to be practical

rational activity in accordance with excellence or the best excellence (*NE*, 1097b33-1098a18). Thus, Brink claims that

"it is in my interest to exercise those capacities that are central to the sort of being I essentially am (*NE* I 7). If I am essentially a person, then a principal ingredient in my welfare must be the exercise of my deliberative capacities." (p. 144)

This argument for the intrinsic merit of the exercising of one's deliberative capacities seems dubious to me, since it still seems to be an open question whether it is good for some particular agent to exercise his deliberative capacities, even though doing so is essential to being a person. Surely the mere fact that one is a person cannot make it good for one to exercise those capacities that are essential to being a person. For in order to reach that conclusion Brink would have to show that it is good for this particular agent to be excellent – or couched in Aristotelian terms, to reach one's human *telos*. Maybe this particular agent's life will be ruined by him trying to achieve excellence – i.e. to exercise his deliberative capacities, and trying to develop them – because doing so would make him utterly unhappy, given his particular aims and goals.[35]

Even if this argument fails however, it still seems as though there is reason, as according to Brink's other argument, which says that people are not deliberatively

[35] This line of critique is similar to the one I launched against Raz's objectivism in section 2.2, that although friendship may be desirable, it is a further question whether this particular agent has reason to pursue it. Sumner (1996: 79) espouses the same objection as I do to Aristotelian teleology. Relying on a distinction between *perfectionist value*, i.e. the sort of value something has if it is a "good instance or specimen of its kind, or that it exemplifies the excellences characteristic of its particular nature" (ibid. 23), on the one hand and prudential value of a life, viz. "how well it is going *for the individual whose life it is*", he argues: "as a conceptual matter the inference from perfectionist to prudential value is never guaranteed; there is always a logically open question. The gap between the two is opened by the agent's own hierarchy of projects and concerns" (ibid. 79).

self-sufficient, that engaging in relationships with intimates is a way of developing one's deliberative capacities, and that developing those capacities is a prudential good.

We have now seen how Brink argues for the rationality of concern for intimates. Intimates, the claim is, extend one's interests, basically because they contribute to the exercise and development of one's deliberative capacities. They do this by making new and superior aims and activities available to me, by contributing to my self-understanding and self-criticism.

One initial worry here is whether such a form of eudaimonism has enough resources to justify the entire scope of morality. So far at least, Brink has merely shown that metaphysical egoism is capable of justifying concern for intimates who are alike and dissimilar to oneself. But how about strangers whom one may not be interacting with at all?

First of all, Brink seems quite optimistic about the possibility of metaphysical egoism explaining how concern can extend to people who are members of the same political association, since there is a commonality of aims, "which is produced by members of the association living together in the right way, in particular, by defining their aims and goals consensually" which in turn "establishes a common good among citizens, each of whom has a share in judging and ruling." (p. 149-150) Hence my concern for other members of the same community is justified on the basis that we as citizens in the same state have certain interests in common, such that I, by allocating some resources to them in effect extend my own interests.

But Brink further suggests that concern, according to his view, can be genuinely universal in scope. Concern is justified on an egoistic basis even to someone – the so-called remotest Mysian – with whom one has had no relation previously, not even indirectly. For

"[t]o the extent that another's actions and mental states are dependent on my assistance, I can view the assistance as making his good a part of my own. Assistance to the remotest Mysian earns me a share, however small, of his happiness, much the way care and nurture of my children grounds posthumous interests I have in their continued well-being." (p. 152)

But how is this concern to be justified on an egoistic basis – surely one may have no possibility of interacting with this remote Mysian, so how could it promote the self-understanding and self-criticism that is a part of one's eudaimonia? It seems clear how this would work in the case of one's own child, for here I have, hopefully, an interest in the happiness my child already; I feel better when I am reassured that he will have a good chance of going on to live a good life.

But in the case of the Mysian how is this argument supposed to run? I cannot experience his happiness, in the same way that I can experience my child's happiness – I might never see him, or know what becomes of him. How can assisting the Mysian contribute to making the agent's as well as the Mysian's life better?

It is quite reasonable to suppose that if I already had some interest in the flourishing of this particular individual (as one normally has in the case of one's child, for instance), then I could view the assistance as making his good a part of my own, because he would now be able to do things which he could not otherwise have done. But this seems to presuppose that one already knows – prior to granting him the assistance – that he will use the goods in a way that is in coherence with one's own interests. But for all one knows the remote Mysian is a terrorist who will use the resources spent on him to harm the giver, or his family, friends, community or state. If so, it seems absurd to claim that the mere act of giving can make the Mysian's good a part of one's own, insofar as his actions and mental states are dependent on one's assistance.

The problem in a nutshell seems to be that there might be some Mysians whose conceptions of the good would make it irrational for others to subsidise them,

because their conception of the good involves eliminating disagreeing views. The upshot seems to be that the mere fact that someone else's good is dependent on my assistance cannot make his good a part of mine – that would seem to depend on the elements of his particular conception of the good.

Perhaps we could amend Brink's claim so that it only applied to people or Mysians whose conception of the good is not too dissimilar from, or at least not contrary to one's own. However, in doing so we run the risk of narrowing our possibility for the self-understanding and self-criticism that comes with being confronted with something different, which is a crucial part of eudaimonia. At least in those cases where the continued existence of the conception of what constitutes a good life is dependent on some kind of support by us, and where we have other reasons – stemming from the fact that this conception of the good essentially involves a hostility towards ours – not to subsidize it, it appears that Brink's theory pulls in two different directions. It seems then that in general some bargain must be struck between extending one's own interests – which would require us to sort the Mysians according to how they agree with our interests, and assist them accordingly, on the one hand – and leaving the various options available in the world, so as to assure a certain amount of diversity – in order to safeguard continued self-understanding and self-criticism, on the other. However, even if this is so, it does seem to follow from Brink's argument that we have a *pro tanto* reason to consider the well-being of even remote Mysians, stemming from concern for our own eudaimonia.

Another reply which Brink might advance would be to simply deny that there is reason to support remote Mysians who have different conceptions of the good than we have. This would dovetail into his objectivist conception of well-being insofar as independent arguments to the effect that the Mysian's conception of what constitutes a good life is objectively wrong (whereas the agent's own is correct) could be produced. If so, one would argue that there is no reason to subsidize the Mysian,

because one would not be developing one's deliberative capacities in an objectively favourable way by making future interacting with such Mysians possible. One does not benefit from having the possibility of continued interaction with ideas about what is good, which do not somehow track the truth about what is good. A terrorist might simply be wrong about his idea of well-being when he, say, kills civilians in order to bring about the downfall of some democratic society, and turn the state into an Islamic theocracy. Thus there could never be any value in having people with such conceptions of the good life around, let alone with such ideas about on the legitimate means to achieving them. On the other hand, if the remote Mysian's conception of endaimonia is objectively correct, then it does seem that he would be extending my interests, which would render it rational for me to support him.

This is certainly a possible line of defence, but one which requires a more substantial discussion than I can offer here. Thus, I shall assume that Brink can successfully answer the objection, and submit it to other criticism in the following section. For now let me offer the following summary of the argument:

(1) One can be, and in fact often is, psychologically connected and continuous with other people (see p. 141),

(2) "Insofar as distinct individuals are psychologically connected and continuous, each can and should view the other as one who extends her own interests" (p. 142).

(2a) persons *qua* persons have a prudential reason to exercise their deliberative capacities.

(2a1) "capacities for practical deliberation – formulating, assessing, revising, choosing, and implementing projects and goals in light of a conception about what is best – are essential to being a person." (p. 136)

(2a2) any given person has a prudential reason to exercise and develop whatever deliberative capacities are essential to being a person (see p. 144).

(2a2.1) "it is in [anyone's] interest to exercise those capacities that are essential to the sort of being [he] essentially [is] (NE I 7). If I am essentially a person, then a principal ingredient in my welfare must be the exercise of my deliberative capacities." (p. 144)

(2a3) since anyone's "persistence as an agent depends upon the extension of [his] deliberative control into the future, […] the exercise of [his] deliberative capacities is part of [his] welfare." (p.144)

(2b) an individual is incapable of achieving a complete deliberative good for herself on her own (see p. 144).

(2c) when people are psychologically connected and continuous, each one is to some degree responsible for the development and exercise of the others' deliberative capacities.

(3) when somebody extends my interests I ought to act according to what other-regarding conduct and concern prescribes towards him.

(C) "interpersonal psychological interaction and dependence provide a metaphysical-egoist justification of other-regarding conduct and concern." (p. 143)

3.3 MORALITY AND AGENT-NEUTRALITY

In the previous section I discussed and summarized Brink's argument. He argued that concern for others is rational on an egoistic construal, because rational concern is not a function of personal identity primarily, but rather of psychological continuity. Concern for one's own eudaimonia requires psychological continuity with others as well as one self, since this in turn turns out to be a way of obtaining a higher degree of self-understanding and hence a way of furthering self-criticism. As we saw, self-

understanding and self-criticism is important because it makes me a better deliberator – and the exercise of my deliberative capacities is a chief part of my eudaimonia, since it enables me to become aware of and pursue superior conceptions of the good. However interesting, I shall eschew the details for the sake of the argument, and consider whether it, if valid, would succeed in answering the normative question.

One objection to Brink's argument states that although it does establish that there is reason to perform moral actions – even towards complete strangers – it does not show that there is overriding or conclusive reason to do so. Although it may be the case that there is some reason to perform the moral act, it is obvious that there may also be reason to perform non-moral acts. For instance, by making a donation to famine relief someone may be extending his own interests (since it earns him a share of the well-being of those who are helped), but it may well be that he does so to a greater extent by using the money to, say, pay for a vacation for him and his family. Here it seems clear that practical deliberation set on extending one's own interests should prefer taking the family on a holiday instead of sending the money away to some faraway Mysian, whom one may never have any contact with or hear from again.

In other cases we can easily imagine how it can be rational for people (either individuals or nations) to try to eliminate other people if their interests cannot be made consistent with their own. At least in some such cases it would seem immoral, and disrespectful, although clearly not irrational, to extend one's interests by attempting to remove obstacles.

However, this consequence of Brink's argument does not show that Brink's argument fails, for he only intended, as he explains, to argue for "the weak rationalist thesis that there is always reason to act on other-regarding demands, such that failure to do so is *pro tanto* irrational." (p. 156) Hence, Brink's argument is compatible

with it being the case that, on balance, there is most reason to perform the non-moral act, but one cannot, on pain of irrationality, simply ignore other-regarding demands.

A further common objection against eudaimonism is that it cheapens the motive because it makes concern for others instrumental to concern for oneself. Morality seems to require not only that we *act* according to moral requirements, but that we do so from the right motive. But according to the picture painted by Brink, agents seem to be motivated solely by concern for their own well-being when they perform morally good acts. They do not appear to benefit the remote Mysian out of genuine concern for him, or because it is their duty to do so, but rather because it is a way of benefiting themselves.

However, as commentators sympathetic to eudaimonism usually point out, the fact that one is pursuing some goal or engaging in an activity or relationship for some final reason R to which the activity or relationship is instrumental, does not generally imply that one is not valuing the goal or activity for its own sake. As J. L. Ackrill explains Aristotle's theory of value as presented in *NE*, 1097a15-b21:

> "In asking what we aim at in action, what its "good" is, Aristotle says that if there is just one end (*telos*) of all action, this will be its good; if more, they will be its good. Now he goes on, there evidently *are* more ends than one, but some are chosen for something else, and so they are not at all *teleia* ("final"). But the best, the highest good, will be something *teleion*. So if only one end is *teleion*, that will be what we are looking for; if more than one are *teleia*, it will be the one that is most *teleion* […].
>
> No reader or listener could be at all clear at this point as to what is meant by "most *teleion*". The word *teleion* has been introduced to separate off ends desired in themselves from ends desired as means to other ends. What is meant by the suggestion that there may be degrees of finality among ends all of which are desired for themselves? Aristotle goes on at once to explain how among ends all of which are final, one end can be more final than another: *A* is more final than *B* if though *B* is sought for its own sake (and hence is indeed a final and not merely intermediate goal) is also sought for the sake of *A*. And that end is more final than any other, final without

qualification […], which is always sought for its own sake and never for the sake of anything else. Such, he continues, is *eudaimonia*: there may be plenty of things (such as pleasure and virtue) that we value for themselves, but yet we say too that we value them for the sake of *eudaimonia*, whereas nobody ever aims at *eudaimonia* for the sake of them (or, in general, for anything other than itself)." (1980: 20-1)

So it does indeed seem compatible with eudaimonism after all that we are concerned for other people for their own sakes; the fact that we treat our own interest in living a good life as the final factor that provides us with reason for action, is, according to Aristotle's theory of value, consistent with us caring for our intimates for themselves.[36]

Brink makes the same point as Aristotle: Dubbing 'final ends' "complete or intrinsic goods" – that is, goods that are chosen for their own sakes – and defining 'eudaimonia' as, the "unconditionally complete" good – i.e. that which is chosen for its own sake and not for the sake of anything else – he explains that

"[i]f the lover treats the good of his beloved as a complete good that is also choiceworthy for the sake of his own *eudaimonia*, the lover is concerned for the lover's own sake while valuing his beloved's well-being for the constitutive contribution this makes to his own *eudaimonia*." (p. 147)

True, there is nothing in the logic of the concepts that prevents one from caring for one's intimate, while also valuing the relationship because of the improvement it makes to the quality of one's own life. None the less, the mere fact that it is *compatible* with Brink's metaphysical egoism that one may – *qua* metaphysical egoist – be concerned about the friend for his own sake does not establish that – as a matter of fact – a metaphysical egoist does care for the intimate for his own sake, or, for that

[36] This in a way should be an unsurprising conclusion, for as I argued above in my discussion of Raz (see section 2.2), friendship need not involve a belief in the mere instrumental value of the friend: One can consistently believe in his final, but agent-relative value.

matter, anyone else. For it is one thing to show that one thing is consistent with another, but something else to show that they indeed are co-existing. Here it clearly need not be the case, for the metaphysical egoist is, after all, an egoist, which is to say, he considers the fact that something results in or constitutes a benefit to him to be the ultimate determinant of the rationality of action. Since there is clearly no necessary connection between him caring for and/or respecting others for their sake and them benefiting him, neither is there a necessary connection between being a metaphysical egoist and being concerned about someone for his sake, whether one stands in a intimate relationship with him or not. What motivates the egoist is the ability of the friend to extend his interest and insofar as, which seems obviously true, the friend fulfilling that function does not necessarily require caring for him for his sake, there is no necessary connection between metaphysical egoism and concern.

However, although the connection between rational egoism and concern for the friend is merely contingent it seems that it would be relatively straightforward to produce an argument within an eudemonistic framework to the effect that there is reason not to be motivated by the consideration which in fact justifies the concern. Just as utilitarians argue that although what justifies an action is the fact that it brings about more happiness than any other alternative course of conduct open to the agent, this ought not to be the agent's motivation since if people were so motivated it will be less likely that humanity will become happy, so the eudaimonist may launch a parallel argument. They may argue that it is inadvisable to be motivated by self-interested considerations, since this is less likely to lead to human flourishing for the person who is so motivated. Being motivated by concern for our loved ones for their sakes without constantly considering whether doing so would contribute to one's own well-being may quite plausibly make one's life prudentially better. In fact, being constantly motivated by self-interest and trying to calculate whether this or that move is instrumental to it may well turn out to be an exhausting and impossible task,

which requires resources which could have been used, say, in the spontaneous enjoyment of the company of one's friends and the unreflected pursuit of one's aims.

If we add such a line of argument to Brink's egoistic justification of morality, then we seem to have been offered a genuine reason to *be* moral. There is reason to be moral because, first of all acting morally means extending one's own interests, and secondly because being motivated by self-interest – i.e. by what provides the reason – is imprudent, because constantly deliberating how best to extend one's interests will make one forgo other pleasures such as appreciating the company of one's friends. One may well be prudentially better off by focusing instead on the good of the friend. But if this is correct then there is reason for one to *be* and not just *act* morally, because being moral, as was argued above, seems to essentially involve acting on a good motive.

Someone who, like Raz, for example, espouses an objective theory of well-being might also argue that concern for the friend for his sake is an essential feature of a *really* valuable friendship, by combining it with Brink's arguments for the good of self-extension such that what he (Raz) would call full friendship is objectively better since it involve a higher level of self-extension. Brink might agree with Raz about this, although of course he would deny what we saw Raz believed, namely that this commits the agent to the agent-neutral value of the friend. Still, it remains open for Brink to reply (to put it in the vocabulary developed in the previous chapter) that it is prudentially better for the agent to be motivated by the *final* value of the friend, without a constant eye to the *instrumental* benefit involved in doing so.

So far I have considered two objections to Brink's justification of morality. The first was that it was unable to justify other-regarding action in cases where self-regarding concern conflicts with other-regarding concern. However as we saw this was entirely consistent with Brink's agenda, for he only intended to argue that there is a *pro tanto* reason to be moral, something which he has certainly argued for persuasively.

Second, I considered the objection that Brink's justification necessarily renders all concern for others purely instrumental to self-concern, such that one is never concerned for the other for his own sake. As we saw, however, it is entirely consistent with Brink's metaphysical egoism that one may be motivated by concern for the other for his sake, although what ultimately provides the reason is self-concern. It even appears possible to argue, on a eudaimonistic foundation, against adopting the reason as one's motivation, which in turn would be an argument for being concerned with the intimate for his sake.

I still believe, though, that Brink's argument fails as a justification for morality, for it seems unable to account for one of the elements which we seem to intuitively recognize as being a part of morality, namely justice.

Justice clearly implies an element of impartiality, which Brink is unable to account for, because he bases all moral reasons on partial considerations. This impartial element in morality however implies the existence of agent-neutral reasons, i.e. reasons which do not involve any essential reference to particular agents. Because Brink justifies morality solely on agent-relative reasons, he is unable to account for this crucial element in morality. Let me try to make this argument more explicit.

Justice requires that we treat individuals as equals, unless they for some reason deserve unequal treatment, and that further, the fact that an individual stands in a special relationship to a particular agent is not in and of itself such a reason. As J. S. Mill puts it in *Utilitarianism*:

"it is, by universal admission, inconsistent with justice to be *partial*; to show favour and preference to one person over another, in matters to which favour and preference do not properly apply." (ch. V, parag. 9)

So justice imposes a requirement of impartiality. For example, if I have advertised a vacant position for a salesman in my business, taking up the point of view of justice

requires that I do not consider the fact that x is my friend and y a complete stranger a reason to prefer hiring x. It would simply be unfair to the stranger, we would say, for me to give preference to the friend just because he is my friend. Instead, I ought to treat them equally, and look at the applicants' qualifications without being biased, and determine who would be most fitted for the job. We can imagine, perhaps, that the applicants are equally qualified or perhaps that the set of qualifications represented by each is incommensurable, in which case e.g. the loyalty one recognizes in the friend might become a consideration. But if so, this consideration is only allowed to tip the balance in virtue of these other features being unable to determine the case.

On Brink's construal however, I ought to pick the friend for the job, because of my own private interest in seeing the friend come along well in life. Therefore, Brink's eudaimonism flies in the face of the requirement of justice, which requires that I treat the two applicants equally. According to Brink's agent-relative conception of morality one would clearly have a weightier self-interested reason to hire the friend than to hire some remote Mysian, whereas justice imposes a requirement of impartiality, which demands that one does not let such a consideration count as a reason in the first place. Viewed from the perspective of justice, there is not only initially equal reason to prefer the two – the fact the one would extend one's personal interests more than another, ought to be no consideration at all. If this is so, then there seems to be at least one agent-neutral element in morality which is left out by Brink's justification. He cannot account for it since he bases morality solely on agent-relativity, which conflicts with it in cases such as the one mentioned. If this is correct, then Brink's argument runs short of justifying morality, since it cannot account for this crucial part of it.

Brink might reply that one's position in a community renders it rational for an agent to be impartial, because if one is not, people might develop hostility towards him,

something which we can easily imagine would impede on the agent's ability to extend his interests. Although this is true as far as it goes, it seems inadvisable to take this route, because it brings Brink dangerously close to the pitfall of 'strategic' egoism. If your being able to extend your interests is conditional on people believing that you abide by the demands of justice, then it would appear that you have no reason to be just (and thus moral) in those cases where you can do so with impunity. If so, then there would still be most reason to act immorally, and help one's intimates first and foremost even when it conflicts with the demands of justice. Thus this reply would contradict Brink's claim that moral demands are counterfactually stable. It follows therefore that Brink's metaphysical egoism does not give us a completely satisfactory reason to be moral.

Brink's problem of course is not peculiar to his justification of morality. On the contrary it seems to be merely an example of a more general theme. By the nature of the case there seem to be two main paths to a justification of moral conduct. The first assumes, as does Kantianism and utilitarianism, as well as Korsgaard and Raz did above, that morality is impartial and therefore about what there is agent-neutral reason to do. Granted this assumption, it has no problems accounting for the impartial elements of morality, whereas it is all the more challenging to allow for the individual ties, commitments, projects and relationships which make human life worth living, because these give us reason to act which are inherently partial. The second approach, exemplified by Brink has the exact opposite problem, for it assumes that morality is about what there is agent-relative reason to do. Accordingly it has no problem explaining the partial element in morality. This has always been a very attractive feature of the Aristotelian approach: it allows due ethical value to those significant personal values, commitments, etc. which seem to most of us necessary elements in a good life, and a key to the development of our potentialities. The difficulty, however, is to account for the impartial element in morality. In Brink's

case it seems to be left completely unaccounted for. Since morality clearly recognizes this agent-neutral element, it would be incorrect to say that Brink has produced an argument to the effect that there is reason to be moral. He fails to account for the element of morality, which consists in a detachment from one's own personal point of view, and which demands equal consideration for the interests of all, unless one has a valid reason for not doing so.

3.4 CONCLUSION

In this chapter I have dismissed Brink's version of eudaimonism – metaphysical egoism – as a satisfactory answer to the normative question. In section 3.1 I explained how Brink's position and argument was related to my general thesis. Brink's argument is an attempt to answer the normative question by showing that concern for one's own eudaimonia gives one a reason to be concerned with other people, or recognize moral demands. Hence in Brink we have yet another attempt to argue someone external to moral demands into accepting them.

In section 3.2 I offered a review and discussion of Brink's ingenious argument for this claim, the main tenet of which was that because we had agent-relative reason to extend our own interests, it is rational for us to be concerned about other people, since this is a means hereto.

In section 3.3., I subjected this view to three different lines of critique. The first of these objected that Brink was unable to justify moral action in those cases where it is preferable, from the point of view of prudence, to perform immoral acts. This objection was found to be out of place, since it was based on the false assumption that Brink took morality to be essentially about what there is agent-neutral reason to do. Since he denied this, he is not caught on this objection.

Second, I discussed the charge that Brink's position made all concern instrumental, such that a metaphysical egoist never is concerned for anyone for their sake.

However, as we saw, Brink's position is first of all entirely consistent with valuing e.g. the friend for his own sake, and second, Brink's argument is easily repaired with an argument to the effect that there is an egoistic reason not to adopt the reason as one's motive.

The third and final objection however showed that Brink's argument in the end was no argument for morality, in the first place. As we saw, justice – one of the most crucial elements of morality – requires impartiality, treating equals as equals. The agent-neutral value justice, however, goes against Brink's purely agent-relative justification of morality, since according to him, people are never equal in the required sense – concern is always a function of self-extension. Hence, although Brink's argument to a large extent is capable of justifying other-regarding concern, it does not give one reason to be moral.

4. CONCLUSION

Over the course of the last three chapters I have considered three attempts to argue, *ad hominem* that there is reason to be moral, even if one does not currently accept the reason-giving force of moral demands. As I pointed out in the introduction, any such attempt must face the challenge constituted by the rational egoist who only recognizes the motivating force of agent-relative reasons. As I further showed – and as I have now illustrated – since a leap from agent-relativity to agent-neutrality cannot be made, one can never succeed in rationally persuading someone into accepting the demands of morality, since being moral in part means being moved by agent-neutral considerations.

One cannot, as we saw Korsgaard attempted in chapter 1, move from agent-relativity to agent-neutrality by assuming that practical reason necessarily operates on the basis of agent-neutral reasons. Her argument, as I revealed, proceeded on the false belief that there are either agent-relative reasons, which are private and hence cannot be shared with others or agent-neutral reasons which are public, and can be so shared. This false premise made her conclude that there are only agent-neutral reasons. However since there are clearly agent-relative values which are shareable (in Korsgaard's sense), the gap remains untouched. Hence she fails to argue someone into accepting moral demands.

In a similar manner, Raz thought that certain sorts of relationships, such as friendships, commits one to the agent-neutral value of the friend (and in turn all people). As we saw in chapter 2 however one need not be committed to the agent-neutral value of the friend, when one believes in the importance and significance of him. Although one might believe in his independent or final value, one might still not believe that others have reason to value him – one may simply believe that he is valuable *to me*. So Raz's argument fails to establish any necessary connection between the prudential value 'friendship' and the agent-neutral part of morality.

The failure of Brink's approach is symptomatic of all attempts to argue for the rationality of moral conduct on an agent-relative basis. His approach actually came quite close to justifying morality, but unsurprisingly failed to justify the agent-neutral considerations which are taken to be reason-giving by normal moral agents. As I illustrated in chapter 3, Brink's metaphysical egoism recommended the immoral act in cases where considerations of justice conflict with our agent-relative, self-interested reason to benefit our intimates.

Hence I conclude – on the basis of my arguments in the introduction and my subsequent discussion of various approaches – that abstract philosophical argument cannot succeed in arguing people into being moral. But that of course does not establish that there is no reason to be moral.

Trying to answer the question "Why should I be moral?", when understood as the project of arguing someone into accepting morality seems to be based on a misapprehension of the nature of morality. Morality, I believe, cannot be derived from first principles, because it (like, it seems, all parts of human conduct) is essentially based in emotional responses which have been instilled into us, through education, socialization, tradition, propaganda, and so on.

If this is a correct view of the origins of normativity, then although there is clearly reason for most of us to do the morally right thing, we should, consistent with the conclusion of this dissertation, not expect that we can argue those who see things differently into believing what we believe if they (are otherwise normal functioning adults who) do not react in the same ways that we do. Moral conduct is not the only rational form of conduct, and because it implies the recognition of agent-neutral reasons, and because the agent-relative/agent-neutral gap cannot be bridged, we cannot argue people in.

BIBLIOGRAPHY

Ackrill, J. L. (1980). "Aristotle on Eudaimonia", in Amélie O. Rorty (ed.). *Essays on Aristotle's Ethics*. Berkeley & Los Angeles.

Aristotle (2002). *Nichomachean Ethics*, translation, introduction, and commentary, by S. Broadie & C. Rowe. Oxford: Oxford University Press.

Brink, David O. (1997). "Self-love and Altruism", *Social Philosophy and Policy*, 14: 122-57.

Foot, Philippa (1972). "Morality as a System of Hypothetical Imperatives", *The Philosophical Review*, 81: 305-316.

Gauthier, David (1991). "Why Contractarianism?", in Peter Vallentyne (ed.). *Contractarianism and Rational Choice*. Cambridge: Cambridge University Press: 15-30.

Gert, Bernard (1979). "Hobbes's Account of Reason", *The Journal of Philosophy*, 76, 10: 559-561.

Heuer, Ulrike (2004). "Raz on Values and Reasons". In Wallace, Pettit, Scheffler, and Smith (eds.). (2004): 129-152.

Hobbes, Thomas (1997). *Leviathan*, Flathman, Richard E. and Johnston, David (eds.). New York: W. W. Norton & Company, Inc.

Irwin, T. H. (1995). "Prudence and Morality in Greek Ethics", *Ethics*, 105, 2: 284-295.

Korsgaard, Christine M. (1983). "Two Distinctions in Goodness", *The Philosophical Review*, 92, 2: 169-195.

(1986). "Skepticism about Practical Reason", *The Journal of Philosophy*, 83, 1: 5-25.

(1996). *The Sources of Normativity*, with G. A. Cohen, Raymond Geuss, Thomas Nagel, Bernard Williams, ed. by Onora O'Neill. Cambridge: Cambridge University Press.

Mackie, John L. (1976). "Sidgwick's Pessimism", *Philosophical Quarterly*, 26, 105: 317-27.

Mill, John S. (1993). *Utilitarianism, On Liberty, Considerations on representative Government*. Geraint Williams (ed.). London: Everyman.

Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.

(1986). *The View from Nowhere*. Oxford: Oxford University Press.

Parfit, Derek. (1984). *Reasons and Persons*. Oxford: Clarendon Press.

Prichard, H. A. (1912). "Does Moral Philosophy Rest on a Mistake", *Mind*, 21, 81: 21-37.

Raz, Joseph (1986). *The Morality of Freedom*. Oxford: Oxford University Press.

(1999). *Engaging Reason*. Oxford: Oxford University Press.

(2001). *Value, Respect, and Attachment*. Cambridge: Cambridge University Press.

Scanlon, Thomas M. (1998). *What we Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press.

Schneewind, J. B. (1977). *Sidgwick's Ethics and Victorian Moral Philosophy*. Oxford: Clarendon Press.

Sidgwick, Henry (1907). *The Methods of Ethics*. 7th ed. London: Macmillan.

Sumner, Wayne (1996). *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.

Wallace, R. Jay (2004). "The Rightness of Acts and Goodness of Lives", in Wallace, Pettit, Scheffler, and Smith (2004): 385-411.

Wallace, Pettit, Scheffler, and Smith (eds.). (2004). *Reason and Value. Themes from the Moral Philosophy of Joseph Raz* (Oxford: Clarendon Press)

Williams, Bernard (1981). "Internal and External Reasons." In his *Moral Luck*. Cambridge: Cambridge University Press.

(1995). "Internal Reasons and the Obscurity of Blame." In his *Making Sense of Humanity*. Cambridge: Cambridge University Press.