Universität Konstanz, Box 216, 78457 Konstanz

To the Executive Committee
of the Zukunftskolleg

Dr Attila Tanyi and Dr Martin Bruder
Zukunftskolleg
Universitätsstraße 10
78464 Konstanz
Tel +49 7531 88-5643
Fax +49 7531 88-4829
martin.bruder@uni-konstanz.de
www.uni-konstanz.de/zukunftskolleg

15.11.2010

Application for Co-Funding for Interdisciplinary Project

Dear Members of the Executive Committee,

Please find attached our joint application for an interdisciplinary project entitled “Overdemanding Consequentialism? Experiments on Moral Intuitions” (project duration: January to June 2011).

We would like to jointly run this project. Our contact details are as follows:

Dr Attila Tanyi
Zukunftskolleg
Department of Philosophy
Universitätsstr. 10
78464 Konstanz

Tel.: 07531-88-5658
attila.tanyi@uni-konstanz.de

Title of Zukunftskolleg project:
Morality and Overdemandingness?
The Case for Authority
(until 30 June 2011)

Dr Martin Bruder
Zukunftskolleg
Department of Psychology
Universitätsstr. 10
78464 Konstanz

Tel.: 07531-88-5643
martin.bruder@uni-konstanz.de

Title of Zukunftskolleg project:
Regulatory Functions of Social Emotions in
Cooperation and Competition
(until 31 October 2011)

The attached application contains the theoretical basis and outline of the project, the work programme, proposed budget, and work schedule.

Yours sincerely,

Attila Tanyi and Martin Bruder

Overdemanding Consequentialism? Experiments on Moral Intuitions

Context and Rationale for Research

Consequentialism is the view that normative properties depend only on consequences. This general approach can be applied at different levels to different normative properties of different kinds of things, but the most prominent example is consequentialism about the *moral rightness* of acts. Consequentialism holds that whether an act is morally right depends only on the consequences of that act or of something related to that act, such as the motive behind the act or a general rule requiring acts of the same kind, as judged from an impersonal perspective. The paradigm case of consequentialism is utilitarianism, whose classic proponents were Jeremy Bentham (1789), John Stuart Mill (1861), and Henry Sidgwick (1907; for predecessors, see Schneewind, 1990.). These classical utilitarians were all *act-consequentialist*: They held that whether an act is morally right or wrong depends only on its consequences (as opposed to the circumstances or the intrinsic nature of the act or anything that happens before the act or anything that relates to the act). They were *utilitarians* because they advocated consequentialism with a welfarist theory of value, that is, a theory that focuses on welfare, well-being, or happiness as the relevant consequence. And since they understood happiness in terms of the balance of the amount of pleasure over pain, they were also *hedonists*. In our research we focus on an objection that has originally targeted these classical utilitarians, but can be employed against any form of act-consequentialism (from now on 'consequentialism').

Let us call this charge the **Overdemandingness Objection** or OD (cf. Williams, 1973; Wolf, 1982; Sidgwick, 1907). OD is built upon two pillars: one, that consequentialism is extremely demanding and, two, that an adequate morality cannot be extremely demanding. Consequentialism requires the agent to promote the good until the point where further efforts would burden the agent as much as they would benefit others. However, the situation that determines what would be best overall is far from ideal: Today's world involves, for example, significant levels of poverty both in the agent's own country and, as a mass phenomenon, in the world as a whole; the levels of charitable donations are insufficient to eradicate it; and the institutions that might make things better are not effective, neither domestically, nor internationally. Given that acting to alleviate poverty is likely to have, in sum, more positive consequences than pursuing individual goals and projects, it seems unavoidable that if one fully accepts consequentialism, one must devote most of one's resources to humanitarian work or to the support of institutions that carry out this work. At the same time, so OD assumes, most people have a firmly held belief that this cannot be right, that people should not be required to sacrifice their life on the altar of morality. This is the second pillar of the objection. If this belief indeed exists, it seems to ground a constraint on admissible moral theories requiring them to avoid unacceptable demands. If they do not, people think, these theories should not be allowed to guide their conduct. OD is an attempt to articulate this constraint.

There are several ways to respond to OD. We can see this if we take a closer look at the structure of the argument above. It is this: (1) consequentialism makes demand D; (2) demand D is intuitively unacceptable; therefore, (3) consequentialism makes intuitively unacceptable demands; (4) if a moral theory makes unacceptable demands, then we have reason to reject it; therefore, (5) we have reason to reject consequentialism. In response, the *strategy of denial* rejects premise (1). It holds that consequentialism does not make high demands, either

because of the empirical circumstances or because of its own internal structure (Cullity, 2004; Hooker, 2000; Mulgan, 2001; Murphy, 2000; Scheffler, 1994). Taking an entirely different stance, the *strategy of extremism* does not deny that consequentialism makes high demands; what it denies is that these high demands are objectionable: that is, it rejects premise (4) (Kagan, 1989; Sobel, 2007; Unger, 1996).

In our research we investigate a third approach that has figured much less, if at all, in the literature due to its empirical underpinnings, which have not yet been established. We examine the truth of premise (2). OD proposes that most people have a firmly held belief concerning the legitimacy of extreme moral demands. In other words, there supposedly exists a **widely shared intuition** that some moral demands are unacceptably extreme. The central goal of this research project is to **empirically investigate** whether common moral intuitions hold that at least some consequentialist demands are overly demanding. That is, the question we want to address is whether common sense morality indeed allows for options in the pursuit of the good, that is, whether it allows us to sometimes pursue the action with suboptimal overall consequences.

In this endeavor, it is important to keep in mind that the criticism articulated by OD can be put in different ways depending on how one understands the charge of overdemandingness. To make the project as philosophically useful as possible, it is advisable to focus on the strongest version of OD. Luckily, from the existing literature, it is clear that there is really only one defensible form of OD. To see this, let us distinguish between three dimensions that the notion of overdemandingness encompasses. The dimension of **scope** refers to the pervasiveness of a moral theory: to the circle of voluntary human action that the theory regards as open to moral assessment. A moral theory is overdemanding in this sense if the circle of actions open to moral assessment is intuitively thought to be too broad. The dimension of **content** deals with the stringency of a moral theory: with the amount of inconsistency that exists between moral directives and the agent's non-moral goals, projects and commitments. A moral theory is overdemanding in this sense if it contains requirements that it should not or if it imposes costs on the agent that are unacceptable, according to people's intuitions. Finally, the dimension of **authority** concerns the inescapability of a moral theory: It concerns the weight of our reasons to act morally as compared to the reasons we have to act non-morally. A moral theory is overdemanding in this sense if it holds actions to be inescapable that people intuitively think they do not have decisive reason to perform.

Since this latter interpretation will be the one that we focus on in our research, a short detour is warranted into the notion of a consequentialist (moral) reason. This is best done by using a distinction between two kinds of reasons (Nagel, 1986; cf. Ridge, 2005). First, non-consequentialists moral reasons are typically **agent-relative**: Our special obligations towards our loved ones (since they are *our* loved ones), certain deontological constraints such as the prohibition on murder (since it would be *me* who commits the crime), and, in particular, our reasons to pursue our own projects and goals (it matters that they are *our* goals), are the prime examples. Consequentialism, on the other hand, claims that moral reasons are **agent-neutral** reasons to promote the good: They are reasons that do not make essential reference to the agent who has the reason (*her* happiness, pleasure etc.). The authority-based reading of OD assumes that these reasons can conflict, and so do we in our research (no doubt, this assumption can be and has been questioned throughout the history of philosophy; but this is a theoretical position, which we will not investigate; for a discussion see Nagel, 1986, and Scheffler, 1992).

Let us turn now to the assessment of the different readings of OD as applied to consequentialism. Based on the literature we suggest that the first two dimensions do not constitute a demandingness challenge. Reducing the *scope* of consequentialism is unwarranted. To mention one thing, moral assessment is context-dependent: Depending on context, any action, no matter how trivial it is, invites moral assessment (Scheffler, 1992). Therefore, if there is a demandingness problem with consequentialism, this is not derived from its unrestricted scope; in fact, unrestricted scope is a desirable feature of moral theories. At the same time, the *content*-based understanding of OD, though popular (the aforementioned strategies to respond to OD all understand the objection in this way), is far from convincing. In particular, there exist arguments that show that focusing on the overdemanding content of consequentialism may cause us to miss our target. A moral requirement may need the backing of reasons in order to make a demand on the agent (Hurley, 2006); and the content-based reading of OD may stem from breaks with consequentialism that are prior to and independent of its demandingness (Sobel, 2007). If either of these arguments succeeds, then, just like pervasiveness, stringency will also not give us an overdemandingness challenge.

These considerations lead to the conclusion that those who advocate OD should follow the authority dimension. Their claim should be that consequentialism is overdemanding because, while being stringent and pervasive, our reasons to meet its requirements override other competing reasons, resulting in situations when it demands us, with decisive force, to do things that we intuitively hold that we do not have decisive reason to do. In our research we will use **experiments** to investigate the **intuitive basis** of this reading of OD.

The present interpretation of OD has two parts. One makes an empirical claim about what people think; the other makes a theoretical claim about what consequentialism demands. Investigating the latter claim would be a purely philosophical exercise; in our research we do not attempt to prove the truth or falsity of this claim. Instead, we will focus on the first, empirical side of OD. That is, we will investigate what people think about the significance of consequentialist requirements and how much weight they attach to these requirements in their every-day deliberations about what to do.

Objectives, Key Propositions, and Basic Experimental Design

We approach this task in the following way. Let us first introduce the so-called **Overridingness Thesis** or OT. According to this thesis, consequentialist reasons override other conflicting reasons of the agent. That is, in a situation where there are several, moral, non-moral, consequentialist and non-consequentialist considerations present, the consequentialist one will win out in the clash of reasons.

The **objective** of our research is to examine the relation between OT and the content of common sense morality (i.e., the set of people's common moral intuitions). In the particular context of the overdemandingness debate, these descriptions of common sense morality matter. They do so in two ways:

- As pointed out above, OD is a constraint on moral theories that is grounded in intuitions. If we find that OT with, at least in part, consequentialist content is part of common sense morality, then OD cannot take the form described above. We cannot say that, intuitively, there is a problem with an overriding consequentialist theory because it demands us, with decisive force, to do things we have no decisive reason to do, when people at the same time intuit that a consequentialist moral theory *should* be overriding.

- There is the issue of how consequentialism can turn out to be overriding in the first place. And one option is by appeal to common-sense morality: that this is exactly what people think about the proper role of morality in their lives and they hold that this morality includes consequentialist requirements. But then if, upon investigation, we find that the kind of morality that people intuitively take to be overriding, is not, at least in part, consequentialist, this path will no longer be open to consequentialism.

Hence, in these two ways, the existing content of common-sense morality has a direct effect on the normative issue of which moral theory to accept.

On the basis of the above implications, we put forward the following three alternative **propositions** concerning the possible ways consequentialism and OT relate to the content of common-sense morality:

1. OT is part of common sense morality, and it has, inclusively or exclusively, consequentialist content.

2. OT is part of common sense morality, and it has exclusively non-consequentialist content.

3. OT is no part of common sense morality.

If Proposition (1) is true, this gives at least some ground for thinking that consequentialism can conform to OT, while making it impossible to run OD; Proposition (2), if true, allows for OD, but deprives us of one way of showing that consequentialism conforms to OT; Proposition (3), if true, also allows for OD, and says nothing about consequentialism, neither positive, nor negative. This is how these admittedly descriptive issues can have a bearing on normative questions in the particular context of our investigation.

In our research we will focus solely on establishing the empirical support for Proposition (1) because it has the most direct relevance for OD: If it receives empirical support, OD fails. The other two propositions have less clear-cut implications.

The next question is how we design our empirical investigation such that it is maximally informative about the truth of Proposition (1). This involves creating paradigmatic situations that can be presented in relatively simple terms to which people can then apply common sense morality without having to recur to philosophical terminology. We suggest that we look for the following moral opinions as indicators of the truth of Proposition (1):

- Experiment 1 will involve conflicting consequentialist and non-consequentialist moral requirements (such as a deontological restriction or special obligation). Proposition (1) would be supported if participants either hold that acting in ways contrary to what consequentialism requires is *immoral* and this *settles* the question of what to do in favor of the moral act (explanation: because they think that an overriding morality is exclusively consequentialist), or they hold that they *cannot* decide because they cannot adjudicate between the competing moral options (explanation: because they think that an overriding morality includes but is not exhausted by consequentialist requirements).

- Experiment 2 will include a clear consequentialist requirement but countervailing non-moral considerations. Proposition (1) would be supported if participants hold that acting in ways contrary to what consequentialism requires is *immoral* and this *settles* the question of what to do in favor of the moral act. This is either because what is immoral

can never be the correct course of action or because, although one can act against a moral requirement, in the present case the countervailing considerations are not strong enough to change the verdict as to the correct course of action.

- Experiment 3 will involve both consequentialist and non-consequentialist moral requirements in addition to non-moral considerations. This experiment is informative because it tests in parallel whether consequentialist reasons override combined non-consequentialist moral requirements and non-moral considerations (which might exert different effects when both are present). Because Proposition (1) can accept that an overriding morality is not exclusively consequentialist, participants can hold *any* opinion when it comes to the comparison between non-consequentialist moral requirements and non-moral considerations (therefore this comparison, although otherwise interesting, will not be specifically addressed in a separate experiment).

Our Material: Moral Intuitions

The investigation of intuitions in moral psychology has become a major focus of contemporary research. But this was not always so. During most of the 20th century, psychologists' treatment of "morality" has focused on the development of reflective and consciously accessible moral reasoning capacities in children. In particular, Jean Piaget (1999/1932) and Lawrence Kohlberg (1984) have introduced influential stage models of moral-cognitive development. However, in the course of the so-called "affective revolution" and under the influence of a new focus on non-conscious processes (Bargh & Chartrand, 1999; Greenwald & Banaji, 1995; Kihlstrom, 1987) the last 25 years have seen a revival of the role of intuitions in theorizing in moral psychology (see Haidt & Kesebir, 2010, for a review).

The concept of moral intuitions reflects the idea that there are moral truths and that people arrive at these truths not primarily by a process of reflection and reasoning but rather by a more immediate process somewhat akin to perception. This is crucial from a philosophical point of view: Intuitions matter for a philosopher because they are taken to have **evidential value** (see Lynch, 2006; Sosa, 2006). Like observations in science, intuitions are the raw data that competing moral theories should at least try to accommodate: If an intuition counts in favor of a theory, this is good for the theory; if an intuition counts against a theory, this is bad for the theory (the theory, philosophers then say, has suffered a counter-example). All this goes only *prima facie*, of course. There can be grounds to discount intuitions, or even not to take them into consideration. It is also possible that, on balance and compared to other theories, a moral theory turns out to be the best available even though it has counter-intuitive implications. Nevertheless, intuitions have initial credibility for (most) philosophers; this is why, unlike in psychology, intuitions were always important in philosophy: they constitute, more or less the only material, philosophers can work on (for a short history of the use of intuitions in philosophy see Appiah, 2008). The subject of our research, OD is a good example. This objection crucially hinges on the content of moral intuitions in that it claims that people intuit that consequentialism is (sometimes) overly demanding and therefore actions demanded by consequentialism are not (always) perceived as morally required.

A possible definition of moral intuitions reflecting this evidential, perception-like role of intuitions is the following:

> "When we refer to *moral intuitions*, we mean strong, stable, immediate moral beliefs. These moral beliefs are *strong* insofar as they are held with confidence and resist counter-evidence (although strong enough counterevidence can sometimes overturn them). They are *stable* in that they are not just temporary whims but last a long time (although there will be times when a person who has a moral intuition does not focus attention on it). They are *immediate* because they do not arise from any process that goes through intermediate steps of conscious reasoning (although the believer is conscious of the resulting moral belief)." (Sinnott-Armstrong, Young, & Cushman, 2010, p. 247, italics in original)

The social psychologist Jonathan Haidt (2001) elaborates on the final characteristic of intuitions – being **immediate** – by stating that "intuition occurs quickly, effortlessly, and automatically, such that the outcome but not the process is accessible to consciousness" (p. 818). Indeed, research by Cushman and colleagues (2006) demonstrated that not all moral principles that affect moral judgments are appealed to by participants when explicitly justifying these judgments. In particular, although participants' judgments reflected that they considered "harm intended as the means to an end as morally worse than harm foreseen as the side effect of an end" (the so-called intention principle), this principle was not invoked when participants justified their judgments. Of course, this finding does not preclude that conscious moral reasoning plays a role in arriving at moral judgments, it suggests, however, that there are at least some influences on moral judgment that operate outside conscious awareness (Greene & Haidt, 2002). Some researchers would hold that moral judgments are fundamentally determined by such intuitions and that elements of reasoning are post hoc rationalizations of such judgments (e.g., Haidt, 2001; but see Paxton & Greene, 2010). Further, moral emotions have been shown to be an important element of moral intuitions (Hofmann & Baumert, 2010).

The immediacy of intuitions is important for philosophers insofar as it ensures that they are **non-inferential**: The moral judgments based on them are not accepted on the ground that they follow from some moral theory or principle (Tersman, 2008). This is essential for them to function as evidence that can, at least *prima facie*, resolve conflict among competing moral theories: they could not support or count against a moral theory were they only to be inferred from that or any other theory.

For the same reason, namely to ensure the evidential value of intuitions, philosophers tend to go beyond the immediacy of intuitions. In the definition quoted above this is reflected in the other two conditions: **strength and stability**. For philosophers these two conditions matter because they help to elevate intuitions to another level: to the level of *considered judgments*, or, as it was recently called, *robust intuitions* (as opposed to the immediate reactions of *surface intuitions*; for the distinction see Kauppinen, 2007). These are those immediate responses of the agent that, so to speak, withstood the test of reflection: They are those (surface) intuitions that a competent speaker would retain under sufficiently ideal conditions such as when the speaker is not biased. The absence of bias, such as prejudice, self-interest, partiality, mistaken heuristics, are seen as essential for ensuring the evidential value of intuitions (Liao, 2008). For the same reason, considered judgments should be based on well-grounded information and sound inference patterns (Rawls, 1971). At the same time, changing focus from the immediate responses of surface intuitions to the more reflective, idealized response of robust intuitions carries the danger of jeopardizing their non-inferential character. For this opens up the possibility of importing one's theoretical convictions, like, for example, inferences from the moral theory one endorses (Doris & Stich, 2005; Liao, 2008).

Methodological Considerations

Our research will focus primarily on moral intuitions understood as immediate responses to an experimental situation and we propose methodological innovations specifically targeted at probing the automaticity and immediacy of intuitions. Although our focus is not on the other two characteristics of intuitions, strength and stability, we will still ask participants several interpretative questions about their responses to the experimental situation to minimize the risk that short-term biases distort participants' intuitions. A more comprehensive test of whether participants' intuitions fulfill these further criteria (i.e., whether they really are 'robust'), would likely need to make use of qualitative methods such as in-depth interviews or focus groups. That is, once we have at hand participants' responses in the experimental situation, we would need to confront them with their moral (surface) intuitions to see whether they withstand reflection. This, however, is beyond the resources available for the proposed pilot project and would rather form part of a more comprehensive follow-up project.

Accordingly, our methodological concerns arise from the non-inferential, non-conscious nature of moral intuitions. In particular, due to their non-conscious nature, people are unable to self-report moral intuitions and such intuitions therefore need to be indirectly inferred. This contributes to a number of methodological challenges.

One important challenge concerns the context in which moral decisions should be observed. The most common way to elicit moral responses is to present participants with verbal moral dilemmas which usually pitch deontological rules (e.g., a prohibition to kill) against widely accepted consequentialist conclusions (e.g., preferring - ceteris paribus - the death of one to the death of many; Waldmann & Dieterich, 2007). The most famous and often-used such dilemma is the so-called trolley problem in which a runaway trolley is about to kill five workers unless the agent decides to hit a switch which will divert the trolley onto another track where it will kill one person. However, there is a question about whether solely including two conflicting moral options in such dilemmas will render them sufficiently self-relevant. To overcome this problem, we will include (in Experiments 2 and 3) non-moral (i.e., "selfish") concerns. This will likely contribute to making the dilemmas more realistic and increase the immediacy of the emotions encountered in such a situation.

A **second** methodological difficulty concerns the identification of the contribution of moral intuitions to the observed moral judgment or behavior. As Figure 1 illustrates, besides consequentialist, deontological, or other moral intuitions, conscious moral reflections as well as non-moral concerns might also play a role in arriving at moral decisions. Drawing on psychological two-system models of cognitive processing that differentiate between a reflective, conscious, rule-based, resource-intensive system on the one hand and an impulsive, non-conscious, association-based, effortless system on the other hand (Sloman, 1996; Strack & Deutsch, 2004) the above definition of moral intuitions makes it clear that intuitions should (at least initially) be processed by the impulsive system. In contrast, the reflective system alone should be responsible for moral reasoning; both systems might participate in the processing of non-moral concerns. Many studies simply infer moral intuitions from moral judgments. In all of our studies, we will introduce a manipulation that will reduce the engagement of the reflective system. In particular, given the limited processing capacity of the reflective system, cognitive load likely reduces reflective processing (Hofmann, Friese, & Strack, 2009). We will therefore introduce a cognitive load manipulation (i.e., keeping a number string in mind; Gilbert & Hixon, 1991) to give us greater confidence in actually measuring moral intuitions rather than reflective processes.

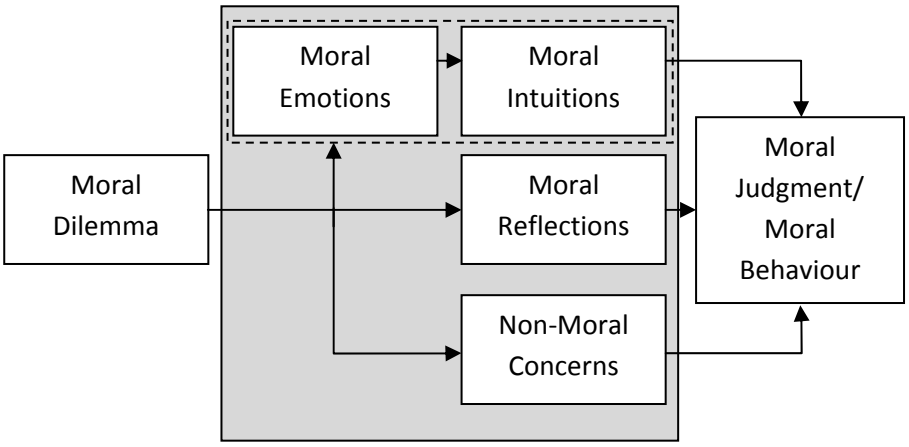

Figure 1. Reactions to moral dilemmas.

A **third** methodological issue concerns the measurement of moral intuitions. Given the automatic, non-conscious nature of the process of forming and applying intuitions, it can be questioned whether they are appropriately measured by verbal scales because these require conscious access to what is being reported. In the domains of attitudes, stereotypes, and self-concept, a broad-range of so-called implicit measures has been developed that aim at providing an alternative assessment of phenomena that may not be consciously accessible (Fazio & Olson, 2003). The most well-known of these measures, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), is based on reaction times in a rapid categorization task with multiple categories. It is interpreted such that category pairings requiring less response time are associated more closely in memory than more difficult and therefore slower category pairings. The IAT has been successfully employed to measure the moral self-concept (Perugini & Leone, 2009) and may also be suitable to measure short-term changes in the activation of specific associations induced by the current context (e.g., Wittenbrink, Judd, & Park, 2001). Using a so-called single target IAT (ST-IAT; Bluemke & Friese, 2008), we will test participants' moral self-concept both before and after the moral decision. We assume that short-term changes in the self-concept are induced by moral intuitions concerning the intermediate decision and that such reaction-time differences therefore allow some inference about whether the chosen options was associated with moral right or wrong.

A **fourth** methodological issue relates to philosophy. Recently, there has been an upsurge in efforts to make a case against intuitions as evidence in moral theorizing. Experiments with philosophical subject matter are used, by the experimenters themselves or by philosophers who interpret the results, to argue that intuitions are not the right material to build and test moral theories on (e.g. Singer, 2005; for a good overview of different attempts see Appiah, 2008; Doris & Stich, 2005). It is suggested that there are no common intuitions, but instead intuitions reflect cultural and socio-economic differences; that intuitions are strongly influenced by the manner in which cases are described or framed; that due to their evolutionary origin, intuitions only embody responses to situations that no longer exist and so on. The common point of these criticisms is that the *origin* of intuitions is such that they cannot serve as evidence for moral judgments: that we can give an explanation of intuitions, which does not assume that the theory for which they are cited as evidence, is true (Tersman, 2008). There are many problems with these arguments (Liao, 2008; Tersman, 2008; Doris & Stich, 2005; Sandberg & Juth, forthcoming), but we are not interested in the pros and cons of 'the case against intuitions'. We see our experimental investigations as supplementing rather than eliminating intuition-based philosophizing.

Work Program

To test Proposition (1) outlined above, we propose the following specific experiments:

Experiment 1: Are Moral Intuitions Consequentialist in Nature?
Overview. The first study will examine how people react to moral dilemmas that involve a conflict between consequentialist moral requirements and other – in this case, deontological – moral considerations. We will primarily assess whether people will perceive the action required by consequentialism to be overriding or not.

Design. The study will follow a 2 (cognitive load: high versus low) × 2 (non-consequentialist demand: high versus low) factorial design with both factors being manipulated between subjects. Participants' implicit moral self-concept will be measured both before and after the experimental task (repeated measures).

Procedure. Participants will complete an ST-IAT on their moral self-concept before and after their moral decision. In the ST-IAT participants will be asked to categorize moral acts (e.g., donation, helping) and non-moral acts (e.g., theft, murder) into the categories *moral* and *immoral*. They will then be asked to also categorize interspersed words related to themselves (e.g., student, German) into each of these categories. Possible differences in the time that these self-related categorizations require (i.e., the degree to which self-related categorizations into the *moral* category are quicker/slower than into the *immoral* category) indicate the strength of the association between the self and morality.

The experimental task will consist of written moral dilemmas that are structurally similar to the trolley problem (for examples, see material provided at http://moral.wjh.harvard.edu/methods/3prin.html). The non-consequentialist, deontological moral requirements will be varied. For example, in the trolley problem, the decision in the condition with low deontological requirements would involve hurting one person and thereby saving five lives; the condition with high deontological requirements would involve killing one person and thereby saving five lives. In each of these conditions, half of the participants will have to decide under high cognitive load (i.e., they have to keep a 7-digit number string in mind) whereas the other half will have to decide under low cognitive load (i.e., keeping only one digit in mind).

We will not only observe people's actual moral decision and the implicit assessment of their moral self-concept, but also ask them about (a) whether each of the possible acts is morally acceptable, (b) whether it is morally permissible, (c) whether it is morally required, (d) whether they would accept a moral theory prescribing the act (i.e., whether such a theory would be overly demanding or not), and (e) what reasons they had for assessing the situation the way they did. The latter question will be posed in an open format and will give us some clues about the stability and strength of participants' moral intuitions.

Statistical Tests. We will examine (a) whether the manipulation of the strength of countervailing deontological requirements causes a shift in moral judgments, (b) whether the manipulation of deontological requirements leads to changed perceptions of the weight of consequentialist moral requirements, (c) whether these effects are moderated by cognitive load, and (d) whether the implicit moral self-concept will be affected by acting in accordance or discordance with consequentialist requirements.

Experiment 2: Do Consequentialist Reasons Override Non-Moral Concerns?
Overview. The second study will examine how people judge and act in situations in which consequentialist requirements clash with self-interest. In this way we will investigate the first scenario in our description of Option (1) above.

Design. The experiment will follow a 2 (cognitive load: high versus low) × 2 (non-moral concern: high versus low) factorial design with both factors being manipulated between subjects. Participants' implicit moral self-concept will be measured both before and after the experimental task (repeated measures).

Procedure. The procedure of Experiment 2 will be highly similar to the one of Experiment 1. The main difference is that countervailing deontological concerns will be replaced by more immediately self-relevant concerns (like, e.g., one's own physical integrity). Again, there will be ST-IATs of participants' moral self-concepts both before and after the experimental task.

Participants will be confronted with verbal moral dilemmas in which consequentialist requirements clash with self-interest. Take, for example, the trolley problem: In the condition with low non-moral concerns, participants will be told that they can stop the trolley but will get hurt themselves. In the condition with high non-moral concerns, they can only stop the trolley at a risk to their own life.

Parallel to Experiment 1, there will be one condition of high and one condition of low cognitive load in both studies.

Statistical Tests. We will examine (a) whether the manipulation of the strength of non-moral concerns causes a shift in moral judgments, (b) whether the manipulation of non-moral concerns leads to changed perceptions of the weight of consequentialist moral requirements, (c) whether these effects are moderated by cognitive load, and (d) whether the implicit moral self-concept will be affected by acting in accordance or discordance with consequentialist requirements.

Experiment 3: Are Consequentialist Reasons Fully Overriding?
Overview. The third study will examine how people judge and act in situations in which consequentialist requirements clash with both non-consequentialist moral concerns and self-interest. In this way we will investigate the third scenario in our description of Option (1) above.

Design. The experiment will follow a 2 (cognitive load: high versus low) × 2 (non-consequentialist demand: high versus low) × 2 (non-moral concern: high versus low) factorial design with all factors being manipulated between subjects. Participants' implicit moral self-concept will be measured both before and after the experimental task (repeated measures).

Procedure. The procedure of Experiment 3 will be highly similar to the one of the previous experiments. However, both countervailing deontological concerns and immediately self-relevant concerns (like, e.g., one's own physical integrity) will be included in the dilemmas. Again, there will be ST-IATs of participants' moral self-concepts both before and after the experimental task.

Participants will be confronted with verbal moral dilemmas in which consequentialist requirements clash with non-consequentialist moral demands and self-interest as described in

the descriptions of the previous studies. Again, there will be one condition of high and one condition of low cognitive load in both studies.

Statistical Tests. We will examine (a) whether the manipulation of the strength of non-moral concerns causes a shift in moral judgments, (b) whether the manipulation of deontological requirements leads to changed perceptions of the weight of consequentialist moral requirements, (c) whether the manipulation of non-moral concerns leads to changed perceptions of the weight of consequentialist moral requirements, (d) whether these effects are moderated by cognitive load, and (e) whether the implicit moral self-concept will be affected by acting in accordance or discordance with consequentialist requirements.

Relations to Current Projects and Interdisciplinarity

The proposed project is closely related to Attila Tanyi's Zukunftskolleg project which discusses OD from a philosophical point of view. It also builds closely on Martin Bruder's research interests in social cognition and affective processes and draws heavily on his methodological expertise in experimental psychology.

The proposed research is therefore highly interdisciplinary in nature. In fact, neither of us could have designed this project without the full support of the other. The project addresses a much-debated issue in moral philosophy and uses current methods from social and cognitive psychology to provide empirical evidence that speaks to that debate. The sophisticated experimental methods (in particular, the manipulation of cognitive load and the reaction time based assessment of the moral self concept) differentiate our work not only from most other work in moral philosophy but also from many contributions in *experimental philosophy*, a new wave of thinking about moral (and non-moral) philosophical questions. At the same time, the focus on a genuinely philosophical debate (i.e., the overdemandingness challenge to consequentialism) differentiates this research from most other empirical studies in the growing field of moral psychology. We hope that this project will be one building block of a stronger bridge between moral philosophy and moral psychology.

However, as for most truly interdisciplinary projects, this is by no means certain. Although each of us can see how this project has high potential to contribute to the own discipline, there is the possibility that many of our disciplinary peers will disagree. Philosophers might be anything but eager to accept constraints on their theories by experiments that – as any methodology – carry their own assumptions. Experimental philosophy has gathered significant momentum recently, with its own blog, journals, and book publications (see here http://experimentalphilosophy.typepad.com); still, it is met with many reservations (see e.g. "the case against intuitions" above). Vice versa, moral psychologists might not be taken by entering into subtle theoretical arguments concerning, say, the dimension of consequentialism that the overdemandingness objection is most promisingly applied to.

Still, in our view, this is the promise that the Zukunftskolleg holds: The opportunity for exchange between researchers from different disciplines and support for the initial stages of high-risk interdisciplinary projects. We will bear most of this risk ourselves, but to keep it in reasonable bounds and not neglect our disciplinary-focused work, we will require substantial support for this project.

Budget

Given the complexity of this project (and associated costs), this project cannot be funded out of our research allowance. Also, given its nature as a pilot project we cannot (yet) turn to larger funding agencies for support because they will require experience in form of preliminary results to grant funding for a project like this.

Experimental Equipment:
The behavioral laboratory in Y123 offers room for three separate work stations. Currently, the laboratory is being equipped with the furniture necessary to install three computers (one of which is already available). Two further computers are therefore needed to make the laboratory fully functional. Besides allowing to conduct the proposed experiments, this will also have long-term benefits for behavioral research in the Zukunftskolleg beyond the current project and would make the Zukunftskolleg research environment more attractive for researchers from disciplines like microeconomics, social and cognitive psychology, experimental philosophy, parts of linguistics, and possibly fields like human behavioral ecology or experimental political science.

-	2 Computers (BW-PC III)	€790,-	
-	2 LG Flatron W2242PK 22" TFT	€390,-	
-	3 Software Licences for MediaLab and DirectRT	€1130,-	($1550,-)
TOTAL		€2310,-	

Participant Recruitment and Payment:

-	3 experiments (45 min and 100 participants each)	€1800,-

Research Support:
Currently, Mr Ramon Gebhard, an M.Sc. student in psychology is employed on a hourly basis. He is open to continuing as a student assistant and preparing his thesis in the area of this grant. However, additional research support will be needed to conduct this project.

-	2 student assistants for 40 hours and 6 months	€4569,-	(€9,52/h)

Dissemination:
One of us will visit the conference "Moral Emotions and Intuitions" in The Hague (May 25 – 27, 2011). This newly established conference reflects the growing interest both in philosophy and psychology into the nature of intuitions and their relationship to morality.

-	Conference visit for 1 person	€750,-

Sum **€ 9429**

Work Schedule

If support is granted, the hiring of research assistants and the setting up of the behavioral laboratory will take place before the official start of the project on 01 January 2011. The experiments will be designed and programmed in January 2011 and will be conducted by April 2011. In parallel, results will be analyzed and the final write-up of the paper will be completed in May and June 2011.

Expected Results and Plans for the Future

We expect the project to produce one peer-reviewed journal publication ideally in a journal at the cross-roads of philosophy and psychology (e.g., *Philosophical Psychology; Journal of*

Theoretical and Philosophical Psychology). Alternative outlets would include either a philosophical journal that has shown interest in experimental philosophy (e.g. *Philosophical Explorations*; *Philosophy and Phenomenological Research*; *Philosophical Studies*) or a social psychology journal open to investigations of moral decision making (e.g., *Journal of Experimental Psychology, Personality and Social Psychology Bulletin, Social Cognition*).

As has been noted above, this proposal outlines a pilot project. The proposed experiments will allow us to gather material (experimental data and a potential publication based on these) to write a larger project application either for the Thyssen foundation or for the John Templeton Foundation, both of which explicitly encourage interdisciplinary project proposals involving philosophy. We believe the project has significant potential for extension (1) by testing the generality of preliminary findings across people (by recruiting representative samples from Germany or other countries); (2) by testing the generality of preliminary findings across situations (by creating experimental situations in which people need to take real rather than fictitious moral decisions); and (3) by testing not only the immediacy, but also the stability and strength of people's relevant intuitions (by including qualitative methodology such as in-depth interviews and focus group discussions).

References

Appiah, Kwame Anthony. 2008. *Experiments in Ethics.* Cambridge, Mass.: Harvard University Press.

Bargh, J. A. and Chartrand, T. L. 1999. 'The Unbearable Automaticity of Being.' *American Psychologist*, 54: 462-79.

Bentham, J. 1961. *An Introduction to the Principles of Morals and Legislation*. Garden City: Doubleday. Originally published in 1789.

Bluemke, M. and Friese, M. 2008. 'Reliability and Validity of the Single-Target IAT (ST-IAT): Assessing Automatic Affect Towards Multiple Attitude Objects.' *European Journal of Social Psychology*, 38: 977-97.

Cullity, Garrett. (2004). *The Moral Demands of Affluence*. Oxford: Clarendon Press.

Cushman, F., Young, L. and Hauser, M. 2006. 'The Role of Conscious Reasoning and Intuition in Moral Judgment: Testing Three Principles of Harm.' *Psychological Science*, 17: 1082-89.

Doris, John M. and Stich, Stephen P. (2005). 'As a Matter of Fact: Empirical Perspectives on Ethics'. In *The Oxford Handbook of Contemporary Philosophy*, eds. Frank Jackson and Michael Smith, 114-152. Oxford: Oxford University Press.

Fazio, R. H. and Olson, M. A. 2003. 'Implicit Measures in Social Cognition Research: Their Meaning and Use.' *Annual Review of Psychology*, 54: 297-327.

Gilbert, D. T. and Hixon, J. G. 1991. 'The Trouble of Thinking: Activation and Application of Stereotypic Beliefs.' *Journal of Personality and Social Psychology*, 60: 509-17.

Greene, Joshua D. and Haidt, J. 2002. 'How (and Where) Does Moral Judgment Work?' *Trends in Cognitive Sciences*, 6: 517-23.

Greenwald, Anthony G. and Banaji, M. R. 1995. 'Implicit Social Cognition: Attitudes, Self-Esteem, and Stereotypes.' *Psychological Review*, 102: 4-27.

Greenwald, Anthony G., McGhee, D. E. and Schwartz, J. L. K. 1998. 'Measuring Individual Differences in Implicit Cognition: The Implicit Association Test.' *Journal of Personality and Social Psychology*, 74: 1464-80.

Haidt, J. 2001. 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.' *Psychological Review*, 108: 814-34.

Haidt, J. and Kesebir, S. 2010. 'Morality.' In *Handbook of social psychology*. Fiske, S. and Gilbert, D. (eds), 797-832, Hobeken, NJ: Wiley.

Hofmann, W. and Baumert, A. 2010. 'Immediate Affect As a Basis for Intuitive Moral Judgement: An Adaptation of the Affect Misattribution Procedure.' *Cognition & Emotion*, 24: 522-35.

Hofmann, W., Friese, M. and Strack, F. 2009. 'Impulse and Self-Control from a Dual-Systems Perspective.' *Perspectives on Psychological Science*, 4: 162-76.

Hooker, Brad. (2000). *Ideal Code, Real World: A Rule Consequentialist Theory of Morality*. Oxford: Clarendon Press.

Hurley, Paul. (2006). 'Does Consequentialism Make Too Many Demands?' *Ethics* 116: 680-706.

Kagan, Shelley. (1989). *The Limits of Morality*. Oxford: Clarendon Press.

Kauppinen, A. (2007). 'The Rise and Fall of Experimental Philosophy.' *Philosophical Explorations* 10: 95-118.

Kihlstrom, J. F. 1987. 'The Cognitive Unconscious.' *Science*, 237: 1445-52.

Kohlberg, Lawrence 1984. *The Psychology of Moral Development : The Nature and Validity of Moral Stages*. San Francisco: Harper & Row.

Liao, S. Matthew. (2008). 'A Defense of Intuitions.' *Philosophical Studies* 140: 247-262.

Lynch, M. P. (2006). 'Trusting Intuitions.' In P. Greenough and M. P. Lynch (eds.), *Truth and Realism*, 227-238. Oxford: Clarendon Press.

Mill, J. S. 1998. *Utilitarianism*, edited with an introduction by Roger Crisp. New York: Oxford University Press. Originally published in 1861.
Mulgan, Tim. (2001). *The Demands of Consequentialism*. Oxford: Clarendon Press.
Murphy, Liam D. (2000). *Moral Demands in Non-Ideal Theory*. Oxford: Oxford University Press.
Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press.
Paxton, JM and Greene, JD 2010. 'Moral Reasoning: Hints and Allegations.' *Topics in Cognitive Science*, 2: 511-27.
Perugini, M. and Leone, L. 2009. 'Implicit Self-Concept and Moral Action.' *Journal of Research in Personality*, 43: 747-54.
Piaget, Jean 1999/1932. *The Moral Judgment of the Child*. Oxon: Routledge.
Railton, Peter. (1984). 'Alienation, Consequentialism, and the Demands of Morality.' *Philosophy and Public Affairs,* 13 (2): 134-71.
Rawls, John. 1971. *A Theory of Justice*. Harvard: Harvard University Press.
Ridge, Michael. (2005). 'Reasons for Action: Agent-Neutral vs. Agent-Relative.' In *Stanford Encyclopaedia of Philosophy*, ed. E. N. Zalta. http://plato.stanford.edu/entries/reasons-agent/. Accessed: 29.10.2010.
Sandberg, J. and Juth, N. (forthcoming). 'Ethics and Intuitions: A Reply to Singer.' *The Journal of Ethics*.
Scheffler, Samuel. (1994). *The Rejection of Consequentialism*. Revised Edition. Oxford: Clarendon Press.
Scheffler, Samuel. 1992. *Human Morality*. Oxford: Oxford University Press.
Schneewind, Jerome. 1990. *Moral Philosophy from Montaigne to Kant*. New York: Cambridge University Press, 2002.
Sidgwick, H. 1907. *The Methods of Ethics*, Seventh Edition. London: Macmillan. First Edition 1874.
Singer, Peter. (2005). 'Ethics and Intuitions.' *The Journal of Ethics* 9: 331-352.
Sinnott-Armstrong, Walter, Young, Liane and Cushman, Fiery 2010. 'Moral Intuitions.' In *The moral psychology handbook*. Doris, John M. (ed), 246-72, New York, NY: Oxford University Press.
Sloman, S. A. 1996. 'The Empirical Case for Two Systems of Reasoning.' *Psychological Bulletin*, 119: 3-22.
Sobel, David. (2007). 'The Impotence of the Demandingness Objection.' *Philosophers' Imprint*, 7 (8).
Sosa, E. (2006). 'Intuitions and Truth.' In P. Greenough and M. P. Lynch (eds.), *Truth and Realism*, 208-226. Oxford: Clarendon Press.
Strack, F. and Deutsch, R. 2004. 'Reflective and Impulsive Determinants of Social Behavior.' *Personality and Social Psychology Review*, 8: 220-47.
Tersman, Folke. (2008). 'The Reliability of Moral Intuitions: A Challenge from Neuroscience.' *Australasian Journal of Philosophy* 86 (3): 389-405.
Unger, Peter. (1996). *Living High and Letting Die*. New York: Oxford University Press.
Waldmann, M. R. and Dieterich, J. H. 2007. 'Throwing a Bomb on a Person Versus Throwing a Person on a Bomb: Intervention Myopia in Moral Intuitions.' *Psychological Science*, 18: 247-53.
Williams, Bernard W. O. 1973a. 'A Critique of Utilitarianism.' In. Smart, J. J. and Williams, Bernard: *Utilitarianism, For and Against*, 77-151, Cambridge: Cambridge University Press.
Wittenbrink, B., Judd, C. M. and Park, B. 2001. 'Spontaneous Prejudice in Context: Variability in Automatically Activated Attitudes.' *Journal of Personality and Social Psychology*, 81: 815-27.
Wolf, Susan. (1982). 'Moral Saints'. *The Journal of Philosophy*, 79 (8): 419-439.