Frontline Learning Research
An official journal of EARLI

Frontline Learning Research Special Issue Vol 8, No. 5 (2020) 70 - 91
ISSN 2295-3159

Explaining Happy Victimizing in Adulthood – A Cognitive and Economic Approach

Gerhard Minnameier[a]

[a] Goethe University Frankfurt am Main, Germany

Article received 1 June 2018 / Article revised 21 November / Accepted 21 November / Available online 1 July 2020

Abstract

While acknowledging the phenomenon of "happy victimizing" (HV), the classical explanation is questioned and challenged. HV is typically explained by a lack of moral motivation (MM) that is thought to develop in late childhood and adolescence. Apart from empirical evidence for widespread HV in adulthood, there are also strong theoretical arguments against the classical explanation. Firstly, there are arguments against the coherence of the very concept of MM. Secondly, while the classical explanation focuses on internal drivers (in the sense of MM), the one proposed in the present paper focuses on the patterns of interaction. Accordingly, HV may depend less on internalised values and individual motivation (whether in terms of moral internalism or moral externalism), and more on the "rules of the game" that are established in social interaction (or not). On this account, HV appears where higher order moral rules are not established and cannot be established, either due the circumstances or due to the unwillingness (or incapability) to play by the rules of these higher order games (where "games" are to be understood in the game-theoretic sense). The ordinary one-shot prisoners' dilemma is a case in point. It precludes promise-giving as well as other higher order moral regimes, but instead forces the agents into a conflict of interest, where everyone has to mind their own business. Moreover, claiming that all players have to pursue their own self-interest, can be understood as moral rule of its own.

Keywords: Happy Victimizer Phenomenon; moral stages; moral reasoning; moral motivation; moral internalism and externalism; game theory; norms and conventions

Gerhard Minnameier, Theodor-W.-Adorno-Platz 4, D-60629 Frankfurt am Main, minnameier@econ.uni-frankfurt.de *DOI: https://doi.org/10.14786/flr.v8i5.381*

1. Introduction

It is well known and empirically well established that 4 to 6 years old children have a marked propensity for the so-called happy victimizer pattern (HVP). This is the combination of violating a known moral rule and feeling good about it (Nunner-Winkler & Sodian 1988; Arsenio, Gold, & Adams, 2006). While older children have more often mixed feelings (bad about acting immorally, good about what they get by doing so), the younger ones do not seem to feel any remorse, even though they know and accept the moral rules they violate.

In the first place, HVP was discovered using a projective method, where the children had to look at a series of pictures that illustrated a case of immoral behaviour. Then they were asked whether the behaviour of the protagonist was ok or not. Finally they were asked how he or she felt and why. Later on, the study was replicated asking the children directly what they would think and feel, if they were in the protagonist's shoes (Keller, Lourenço, Malti, & Saalbach, 2003). Even in these circumstances, about 50 per cent of the participants showed HVP.

The proposed explanation was straightforward. Since children obviously understand and seem to have internalised the moral norms, they are said to have the *moral knowledge* necessary for moral action, but lack *moral motivation* (Nunner-Winkler & Sodian, 1988; Nunner-Winkler 2007; 2013; Arsenio et al., 2006; Krettenauer, Malti & Sokol, 2008). Moral motivation has even been thought to be the crucial ingredient that developed only gradually in late childhood and adolescence and prevented many from doing the right thing (Nunner-Winkler 1999; 2007; Malti & Krettenauer, 2013).

However, what *is* moral motivation? Even though the concept is old, it is still difficult to grasp (Brink, 1997; Smith, 1994/2005; Zangwill, 2003; Minnameier, 2010; Malti & Krettenauer, 2013; Wren 2013; Heinrichs, Minnameier, Gutzwiller-Helfenfinger, & Latzko, 2015; Rosati, 2016). Furthermore, concerning its development, the proponents of moral motivation remain mostly silent. While Nunner-Winkler has thought that it develops gradually throughout childhood and adolescence (see above), others think it is there right from the start (Malti & Krettenauer, 2013, referring to Warneken & Tomasello, 2009, who have found strong evidence for pro-social behaviour among toddlers).

What we know, however, is that HVP has turned out to be salient even among young adults and contrary to the classical explanation, according to which HVP should vanish in late childhood (Nunner-Winkler 2007; Krettenauer, Malti & Sokol, 2008; Minnameier & Schmidt, 2013; Heinrichs et al. 2015). Beyond the narrow frame of research on HVP, there is also huge evidence for happy-go-lucky cheating and other forms of moral victimisation among adults (Batson et al., 1999; 2002; Ariely, 2012, Rustichini, & Villeval, 2014).

In particular, it is well known that students of economics and business administration act more selfishly in various experiments than students of other subjects (Marwell & Ames, 1981; Frey, 1986; Carter & Irons, 1991; Frank et al., 1993; 1996; Frey & Meier, 2003; Rubinstein 2006). Further analyses have mainly focused on whether this points to a self-selection of self-interested individuals into economics-related studies or to an "indoctrination effect", which means that economics is taught in such a way that makes students more selfish and competitive. The differentiation was introduced by Carter and Irons (1991) who found evidence in favour of the self-selection hypothesis, based on the ultimatum game. However, Selten and Ockenfels (1998) as well as Frank et al. (1993) found evidence for an indoctrination effect.

Within this economic body of research, however, it has never been asked how people actually take their decisions and what is really appropriate in specific circumstances. After all, we know that people use moral principles in a situation-specific manner (Krebs & Denton, 2005; Rai & Fiske, 2011; Minnameier, Beck, Heinrichs, & Parche-Kawik, 1999; Minnameier & Schmidt, 2013). Hence, the question arises how economists and non-economists actually take their decisions and what is (more or less) appropriate. By the same token, HVP in general might be explainable as an action pattern that is morally justified (or at least justifiable) under specific conditions.

Thus, while the phenomena are almost crystal clear, the explanation is not. In section 2.1, I will summarise arguments against the received explanation from moral motivation that I have elaborated elsewhere in detail (Minnameier 2010; see also 2012; 2013). On this account, the common understanding of moral motivation has to be turned around almost completely. What I suggest instead in the remainder of section 2 is to replace the false dichotomy of moral cognition and moral motivation by a more comprehensive theory of moral judgement and agency that includes processes of abduction, deduction and induction. This approach converges not only with the integrative account of Rai and Fiske (2011), but also with a "reason-based theory of rational choice" (Dietrich & List, 2013a and b) that allows us to integrate cognitive moral psychology with rational choice theory (section 2.2). As it turns out, however, this theory of moral decision-taking has to be integrated into a game-theoretic view, because moral principles represent more than just personal values (section 2.3). In section 3 a study is presented that contains important evidence in favour of the reason-based approach, and section 4 contains an extensive discussion that also introduces further theoretical ramifications. Section 5 concludes.

2. Moral Motivation and Beyond

2.1. The Problem of Moral Motivation

The idea of moral motivation has different sources. For instance, Kant needed it to explain why an individual ought not only to follow moral principles, but to engage in moral reasoning and develop moral principles in the first place (Ameriks, 2006). This is why Kant says that "(t)here is nothing it is possible to think of anywhere in the world, or indeed anything at all outside it, that can be held to be good without limitation, excepting only a *good will*" (2002/1785, p. 9 [4.393][1]). In modern philosophy and psychology, the concept has been used to explain HVP-like behaviour (in terms of a lack of moral motivation). Philippa Foot (1972) kicked off the modern debate in philosophy observing that we can very well be indifferent to morality without being irrational (Zangwill, 2003). And in moral psychology it was Augusto Blasi (1984) and James Rest (1984) who held against Kohlberg that moral reasons did not motivate moral action directly, but needed to be sided by judgements of responsibility (Blasi) or moral motivation (Rest).

Both strands of reasoning, the one in moral philosophy and the one in moral psychology, are directed against a view traditionally ascribed to Socrates. Socrates is said to have claimed that knowing what morality (or virtue) demands motivates virtuous action and that acting otherwise indicates the agent's ignorance about the moral course of action (see e.g., Brickhouse & Smith, 2010, esp. chap. 3).[2] Advocates of *moral externalism* hold that one can very well know what the moral course of action would be, yet not be motivated and therefore act in a different way. Moreover, they do not interpret this as a mere case of weakness of the will, in which agents would act against their own intentions and possibly be racked by remorse as a consequence. Their view is that a moral judgement has to combine with a desire, where both are related only contingently. Therefore, on the externalist account, moral judgement does not motivate by itself, but needs to be seconded by moral motivation (see e.g., Rosati, 2016).

This is precisely the way in which James Rest, in the psychological camp, has defined moral motivation, i.e., to "select among competing value outcomes of ideals the one to act on; deciding whether or not to try to fulfil one's moral ideal" (1984, 27). And he differentiates it from willpower, which means "to execute and implement what one intends to do" (ibid.). This is, by and large, the current

[1] As a world-wide citation standard, the two numbers refer to the volume and the page in the famous "Akademie-Textausgabe".

[2] Brickhouse and Smith show that, contrary to the received view, Socratic moral psychology is not naïvely cognitivistic, but more in line with what it is claimed in the present contribution.

state of affairs in moral psychology concerning the concept of moral motivation (see e.g. Thoma & Bebeau, 2013; Nunner-Winkler, 2013).

Against this view, I have argued elsewhere that this conceptual framework is inconsistent for two reasons (see Minnameier, 2010; 2013). The first is that Rest's definition of moral motivation implies a judgment: If individuals are not driven against their will (which would indicate lack of willpower rather than lack of moral motivation), but *select freely* among competing value outcomes, this kind of selection requires a decision based on some criterion. In other words, moral motivation would not be independent from moral judgement, but would have to involve some kind of moral judgement resulting in an intention that motivates action. In fact, deciding whether to go for some personal benefit or to dispense with it for some other-regarding motive is even *the* paradigmatic case of a moral problem as such. Therefore, the classical definition of moral motivation seems ill-conceived?

The second point pertains to what makes *moral motivation* moral. How could moral motivation possibly be distinguished from other kinds of motivation, if not for an underlying reason or so? I see no other way to explain the moral aspect of moral motivation than to refer to some underlying reason. Hence, moral motivation would have to be derived from the conviction of being morally obliged in some way.

2.2. Inferential Moral Reasoning and a Reason-based Theory of Rational Choice

On the cognitive view, *moral motivation* in the sense of Rest's third component is not just eliminated, but rather replaced by a specific kind of judgement within a more comprehensive notion of moral judgement. This conception allows for three different parts of moral reasoning which play distinctive roles in the formation of a moral intention. In particular, this broader notion of moral reasoning comprises abduction, deduction, and induction, which in turn are based on the Charles S. Peirce's pragmatist theory of inferential reasoning (see Minnameier, 2004; 2017).

According to this approach, both the adoption and the application of moral principles are mediated by three characteristic inferences called "abduction", "deduction" and "induction". Any kind of development is triggered by a negative, i.e., disconfirming, *induction*, which means that a certain principle that one previously adhered to fails to do the job in the given situation. For instance, if a child uses the Golden Rule ("Do unto others as you would have others do unto you") and now faces a decision at school about where to go for a class outing, this rule might not suffice. For even if everybody followed this principle, no unitary decision might be obtained. Hence, instead of trying to square diverging *individual interests*, a decision has to be taken at the level of the *group*. Majority voting would be a solution for this kind of problem, and this solution allows us to transcend individual interests in order to determine what is best for the group as a whole. Moreover, group decisions call for loyalty on the part of the outvoted members.

The shift onto the higher stage can be understood as being mediated, firstly, by a *negative induction* which leads to the insight that the Golden Rule *fails* to solve the problem at hand.[3] Secondly, a new principle has to be invented, which is captured by an *abductive* inference. Thirdly, *deduction* tells us what follows, if the principle is applied to the situation at hand, in particular what kind of action would have to be taken. Fourth, and finally, *induction* leads to the adoption (or rejection) of that principle as a guideline for the present situation, but equally for all situations of the same type.

These inferential processes can also be assumed to mediate moral action, where moral principles are not invented *ab ovo* in a genuine developmental transformation, but merely activated in relevant situations (see Minnameier, 2013). In those cases, the inferential reasoning may take an explicit or

[3] This is an induction, because it is not just a formal deductive judgement based on deductive premises, but determines a belief about what is feasible and effective in real world interactions. Furthermore, this belief includes the insight that the Golden Rule not only fails in the present situation, but would equally fail in all situations of the same kind as the present one. In this sense, any induction, even if it actually only addresses a single situation, is always generalizable, in principle. And this applies to positive (confirming) as well as negative (disconfirming) inductions.

implicit (habitual) form, especially in well-rehearsed situations, for which habitual action schemes have been established. In the latter case, we may immediately know how to react without engaging in any (further) reasoning, but would have to be able to explicate our reasons on demand.

The inferential approach meshes perfectly with a reason-based theory of rational choice (RBT) (Dietrich & List, 2013a and b; Minnameier 2016a). Dietrich and List strongly criticise the behaviourist approach that is still prominent in economics today (see e.g. Gul & Pesendorfer, 2008) and which relies on the concept of "revealed preferences" (Samuelson, 1938; 1948). This means the preferences are revealed from choice data and rationality assumptions like completeness and transitivity, just to name the most important. If you choose apples, where you could have had pears for the same price, this reveals that you prefer apples over pears. However, our "real" preferences are usually of a deeper psychologic nature, so that we might buy a bunch of red roses to impress a beloved person, not because we find them particularly attractive and worth the price for what they are as such.

Therefore, Dietrich and List formulate a rational choice theory in which motivating reasons (i.e., the underlying drivers which need not necessarily be made conscious) play a central role. We could also call them the "fundamental preferences" that express what we are really striving for, as opposed to the concrete or instrumental preferences, like the one for red roses in the example. The latter are formed based on the former and the specific restrictions that hold in a given situation (like time, money, and so forth).

If we take moral principles as fundamental preferences or "reasons" in terms of RBT, then the situation-specific adaptation of morality is theoretically tractable within a rational-choice-theoretic context. The situational cues provide information about the positive and negative restrictions (affordances and constraints), which we would have to understand as "beliefs", because what counts for the explanation of behaviour is how the agent views the situation, not an objective account of it. The fundamental preferences and the beliefs determine the concrete preferences that we could also call "intentions" in the context of morality. This in turn, would determine a rational action. However, people might fail to act according to their own intentions, especially for lack of willpower, which allows us to include "irrational choice" as well.[4]

In the inferential context, we can say, accordingly, that a moral problem perceived as a feature of the situation *abductively* leads to a moral principle that captures this problem. From this principle and the situational premises we can deductively derive one or several moral courses of action, which just follow from applying the principle to the situation. The last, *inductive*, step is to determine whether this solution of the moral problem can be appropriately implemented. Sometimes, the price one has to pay, or the risk, might be too high, so that we might legitimately discard a certain course of action at this stage. However, if there are no such problems, we commit ourselves and so a moral intention is formed. This would then have to be carried out on pain of irrationality.

Finally, I think this reconstruction is valid also for intuitive moral agency, because if we do not assume any kind of moral reason or principle to underlie the act, we could never distinguish self-interested action from moral action. Therefore, even an intuitive helping behaviour that ought to be characterised as "moral" must imply a moral point of view (e.g., that the other person is in need of help and therefore should be helped). Consequently, some kind of moral cognition would always have to be abducted, even in non-deliberative moral decision making (see Minnameier, 2016b; 2017; Hermkes, 2016).

[4] Note that already in Gary S. Becker's economic approach to human behaviour (1976; 1993), preferences are conceived as very fundamental. However, he uses Rational Choice Theory to explain any kind of (stable) human behaviour, i.e., he uses the principle of rationality as an explanatory tool, so that clearly nothing remains as irrational, since what is irrational on this account, is simply not rationalisable. Weakness of will, however, is part of an explanation, and then this weakness is interpreted as a constraint on the agent's actions, which rationalises even choices that run counter to the agent's intentions.

2.3. Morality in the game-theoretical context

Rational choice theory branches out into decision theory and game theory (see e.g., Binmore, 2009, p. 25). The former relates to the choices a rational agent takes in a certain environment ("games against nature"), whereas the latter concerns the interaction with one or more other rational agents. On the reason-based account, moral agency is modelled within a decision-theoretic framework, and moral values are treated as personal values (fundamental preferences) according to which the individual wants to live. However, from such a point of view, moral judgements are (involuntarily) reduced to prudential judgements, and so is moral action reduced to prudential action, because whatever an agent does, it is always explained in terms of maximising utility with respect to satisfying fundamental (but still personal) preferences (whether we call them morals, values or virtues).

This raises the question of what morality really is. This question is very topical also in the context of prosocial behaviour among toddlers and apes (for reviews see Paulus, 2014; Killen & Smetana, 2015). It appears as if they have a sense of morality or justice. However, it may be a mistake to infer directly from prosociality to morality. First of all, agents may have altruistic or other-regarding orientations, but these would none the less be *their own* orientations. They might be explained in terms of vicarious experience caused mirror-neuron activity. However, in whichever way these orientations are explained, the hard problem remains. The hard problem is that stretching morality to include mere prosociality would make the distinction between *morality* and *prudence* obsolete or impossible. Conversely, if (first person) prosociality is distinguished from (third person) morality, the morality requires a third-person perspective (or meta-perspective) from which the differentiated perspectives of self and other(s) are looked upon and coordinated in some way.

This idea of a hard problem can even be taken further. We typically conceive morality in terms of (other-regarding) personal values. People who have internalised such values are regarded as highly moral people in the sense that they are intrinsically motivated to act morally. Conversely, those who are only extrinsically motivated appear to use morality instrumentally and are therefore considered not to be morally motivated. In other words, they are not moral agents, but only do as if they were moral for some selfish reasons. Again, they seem to be guided by a prudential rationality rather than a moral rationality. However, what appears as truly moral on this account, still runs into the hard problem, because if agents follow their personal values and try to maximise utility in this respect, they will always have to follow a prudential rationality, whatever the specific content of their values and however other-regarding they may be.

This hard problem remains, because morality is reduced to a "decision-theoretic" problem rather than a "game-theoretic" one. However, morality is basically a social project, not a question of the good life for the individual, and hence not just a question of personal values. If moral principles are to make any sense, they do not only address the agent as an individual, but all agents involved in a moral problem. Moral rules are social rules, and these rules have to be accepted and heeded by the set of agents to whom the rules apply.

A further common misunderstanding or misconception is this: If we think of morality in the game-theoretic sense, we often treat moral issues in terms of a *zero-sum game*. A zero-sum game is one in which a fixed sum of money or amount of goods is to be allocated. The overall sum does neither increase nor decrease. In this very sense we think that rich people should transfer some of their riches to the poor, thereby leaving the total unchanged. On this account, those who give suffer a loss (for moral reasons), and the needy enjoy the benefit.

However, apart from zero-sum games we have two other types, i.e., *coordination games* and *cooperation games*. Coordination games are less problematic, because coordination is always beneficial for every agent. For instance, in some countries people drive on the right-hand side, in some they drive on the left-hand side. None the less there is no reason, say, for those from the continent to drive on the left while in the United Kingdom, and those from the UK would have no incentive to drive on the right while on the Continent (unless they wanted to commit suicide). Where we drive is a question of “conventions”, because one just has to follow a uniform rule, and there seems to be no deeper reason, at least no moral one, why we should decide to drive on the right or on the left. Thus, *conventions* are the solution concept for coordination games, and I think domain theorists like Turiel (2002) and Nucci (2008) could benefit from consulting game theory to distinguish conventional problems from moral problems.

Cooperation games are different. The prisoners’ dilemma is a classic example. In this type of situation, “cooperation” would benefit everyone (like in a coordination game), but it does not constitute a stable equilibrium in the sense of a Nash equilibrium. In the prisoners’ dilemma the agents have an incentive to defect, which takes them to a social trap in the form of a Pareto-inferior Nash equilibrium. Their only way to overcome this deadlock is to invent a so-called institution, i.e., a rule backed by suitable sanctioning mechanisms that ensures cooperation. If the prisoners’ dilemma were not as restrictive as it is (where the agents are not allowed to communicate with each other), they would perhaps agree not to confess to the judge and threaten each other with the punishment of comrades outside the prison or with their own fierce revenge in case the other might defect.

However, such an institution is tantamount to a moral rule, in particular that of a mutual promise or contract. And the potential sanctions may not only be negative, but also positive (respect for one’s reliability and a mutual willingness to cooperate in the future). Hence, moral principles can be straightforwardly reconstructed as solution concepts for so-called cooperation problems? I think this hits the point, and I also think that both the decision-theoretic interpretation as well as the understanding of morality in terms of a zero-sum logic are severely flawed. These misconceptions are at the heart of fatal errors about moral motivation and the related normative questions (see e.g. Minnameier, 2013, and the discussion of empirical results below).

3. Empirical Evidence of HVP in the Light of the Reason-based Approach

3.1. Review of a Study on Morality in the Economic Context

Here, I would like to summarise the data and results of a recent study on how economists and non-economists choose to act in different framings of the famous prisoners’ dilemma (Minnameier, Heinrichs, & Kirschbaum, 2016), which also yields important insights about how prevalent HVP is among different groups and in different contexts, and on how it might be brought about. This is essential for the crucial question of whether those who show the pattern are really motivated immorally or not, and whether the self-interested behaviour is to be morally condemned and educationally tackled or is rather acceptable or even desirable as an adequate situation-specific response to the constraints the individual faces.

As mentioned in the introduction, there is clear evidence that students of economics and business administration behave more selfishly than others, at least in certain situations. And by the same token we may assume that HVP is more prevalent among these students. However, not so much is known about the motivation of either self-interested or other-regarding behaviour. Moral psychologists tend to ask why people act selfishly and fail to follow moral rules. Economists wonder about other-regarding behaviour which they consider irrational. Thus, there seems to be a need for clarification.

The prisoners’ dilemma is a case in point. It has been used to measure moral motivation (see Nunner-Winkler, 2007), where cooperation indicates high moral motivation and defection low moral motivation. On this view, however, the moral course of action is the exact opposite of rational action in terms of an economic analysis. Conversely, defection, which is the dominant strategy and thus rational from an economic point of view, is thought to indicate a lack of moral motivation. Hence, you could either be “rational” or “moral”, but not both.

Luckily, the reason-based approach sketched above allows us to reconcile morality and rationality. Based on it, we may assume that even though economists might generally be more self-interested than students of other subject matters (especially following the self-selection hypothesis), not all those who take self-interested decisions may do it for lack of moral motivation, but may simply be more realistic and prudent in their decision-making. To test this assumption, we have carried out a study using the prisoners' dilemma (PD) in the two framings already discussed above, i.e., we called it "Wall Street Game" in one condition and "Community Game" in the other. Additionally, we let participants express their feelings and reasons. Our hypotheses in this context were the following (we use "economists" as a short cut for "students of economics and business administration", and "B&E education" for "business and economics education"):

1. Compared with other students, economists (a) defect to a greater extent and (b) are less vulnerable to the framing (since they know they are in a PD and that defection is the dominant strategy).
2. Accordingly, we expect (a) more happy victimizers than happy moralists among economists and (b) fewer happy victimizers than happy moralists among the students of other subjects.
3. We expect that at least in part these differences do not indicate differences in moral judgement and/or moral motivation, but merely different views of the situational constraints.
4. Since we have also students of B&E education in our sample, we expect them to score at intermediate levels as compared with the other two groups.

3.2. Sample and design

The sample consists of 481 undergraduates from two German universities (Frankfurt am Main and Bamberg).[5] 54 percent of the sample are bachelor students of economics and business administration ("economists"), 14 percent are bachelor students of B&E education, and 32 percent are teacher students with no economics-related subject (henceforth "teacher students"). 39 percent of our participants are male, 61 percent are female. The average age is 22.4 years.[6]

The participants were presented a classical PD in two different frames. For one half it was called "Wall Street Game" (WG), for the other half it was called "Community Game" (CG) (see Liberman et al., 2004; Ellingsen et al., 2012). They were randomly assigned to one of the two conditions. Of the 481 participants, 227 were in the CG-condition, 254 in the WG-condition (χ^2 = 1.516, p = 0.218). The description and instructions were exactly the same in both conditions with the only exception that the respective name of the game appeared three times in the instructions. They had to decide between two options, A and B, and were given the following information:

- If both you and the other person choose A you both get €50.
- If both you and the other person choose B you both get €20.
- If you choose A and the other person chooses B you get €5 and the other person gets €80.
- If you choose B and the other person chooses A you get €80 and the other person gets €5.

After having chosen one of the two possible strategies – which are commonly called "cooperate" (A) and "defect" (B) – they had to explain, first, why they decided the way they did. After this, they had

[5] We have a total of 587 participants in the study. However, 78 of them study something other than the three groups discussed here of have failed to specify the study programme. 28 have not completed the questionnaire and were therefore excluded from the analysis.

[6] While the proportion of male and female participants is fairly equilibrated for economists (50/50 percent) and B&E education (42/58 percent), it is not for teacher students (21/79 percent). This study does not focus on the gender aspect. None the less, we control for gender in the subsequent analyses of moral orientations. In terms of age the differences are small (economists: 22.42 years; B&E education: 24.83; teacher students: 21.21 years) but significant (according to the Kruskal-Wallis-test, since variances are heterogeneous).

to rate how they feel about the decision on a four-point Likert-scale (good, rather good, rather bad, bad) and, finally, they had to explain their feelings.

The data on the participants' emotions allowed us to code the answers in the Happy-Victimizer framework (Arsenio, Gold & Adams, 2006; Nunner-Winkler, 2007; 2013), where good or rather good feelings indicate "happiness" and bad or rather bad feelings indicate "unhappiness". Concerning the choice options, cooperation codes for "moralists" and defection for "victimizers". Thus we classify the answers as happy (HV) or unhappy victimizers (UV) and happy (HM) or unhappy moralists (UM).

3.3 Results

As for the framing in terms of WG and CG, no framing effect can be identified in the total sample ($\chi^2 = 0.112$, *df* = 1, *n* = 481, *p* = 0.783). The same is true for economists ($\chi^2 = 0.007$, *df* = 1, *p* = 1) and B&E education ($\chi^2 = 0.122$, *df* = 1, *p* = 0.804), but a rather strong one among the non-economist teacher students ($\chi^2 = 5.374$, *df* = 1, *p* = 0.029), with 80 percent cooperation in the CG-condition and only 63 percent in the WG-condition (see Table 1). Economists only cooperate by 45 percent (CG) and 44 percent (WG) respectively. This is not surprising, since economists know the prisoners' dilemma and have identified it as such. Apart from the framing, however, the defection rate of economists is much higher, in general, than that of non-economists. Students of B&E education score at intermediate levels.

Table 1
Framing, Cooperation and Defection in the PD

Study Program	Framing	Defection	Cooperation	
Economists	CG	55.3%	44.7%	} n.s.
	WG	55.8%	44.2%	
B&E education	CG	44.8%	55.2%	} n.s.
	WG	40.5%	59.5%	
Teacher students	CG	19.3%	80.7%	} p = .029
	WG	37.1%	62.9%	
Total	CG	44.9%	55.1%	} n.s.
	WG	46.5%	53.5%	

As Table 1 also reveals, the study programme has a significant effect on the decisions taken (χ^2 = 24.799, *df* = 2, *n* = 476, *p* = 0.000). This effect remains, if we control for gender (male: χ^2 = 10.988, *df* = 2, *n* = 187, *p* = 0.002 (one-tailed); female: χ^2 = 13.208, *df* = 2, *n* = 289, $p < 0.001$, one-tailed).[7] Thus, hypotheses 1a and 1b are both strongly confirmed. And concerning hypotheses 4 we can state that students of B&E education are at an intermediate level with respect to the proportion of cooperation and defection.

Perhaps even more importantly, we see a major difference in terms of agency (HV/UV/HM/UM), with 55 percent HVs among economists, but only 31 percent among non-economists. Conversely, 67 percent of the non-economists are HMs, whereas only 36 percent of the economists are HMs (see Figure 1; χ^2 = 26.094, *df* = 6, *n* = 337, *p* = 0.000, based on Fisher's exact test). If we control for gender, results are still statistically significant (male: χ^2 = 11.016, *df* = 6, *n* = 130, *p* = 0.044 (one-tailed); female: χ^2 = 12.052, *df* = 6, *n* = 204, *p* = 0.026, one-tailed). Hypothesis 2 is confirmed by these data, and so is hypothesis 4.

[7] In the analyses on gender differences, only 476 cases are included, because the remaining five have failed to indicate their gender.

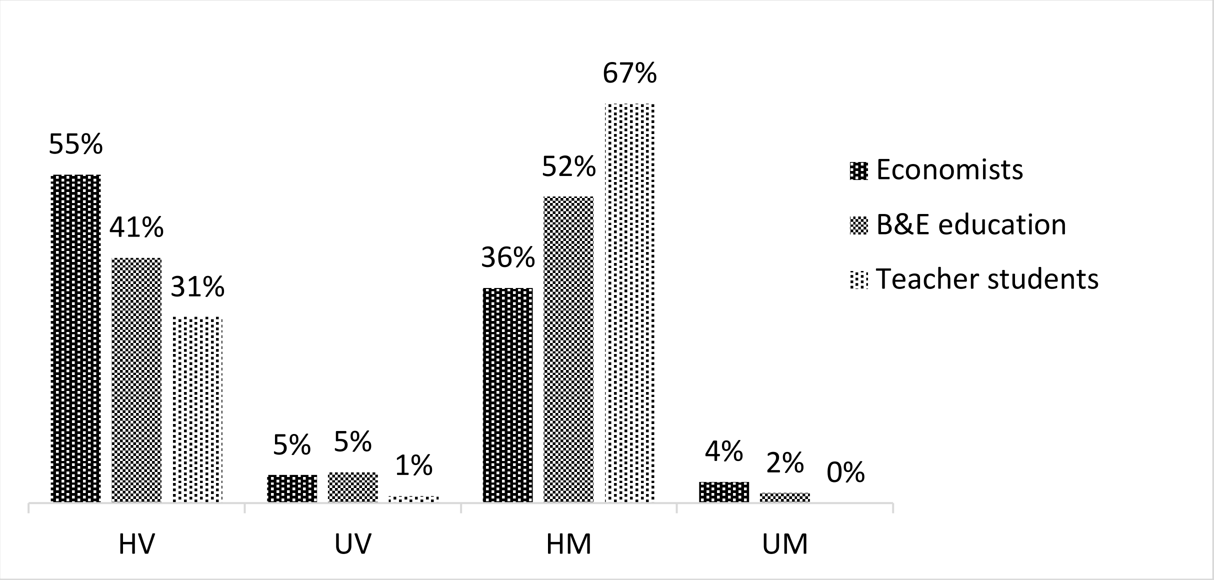

Figure 1: The Proportions of HV, UV, HM, and UM.

Another important result relates to the reasons the participants give for their choices and feelings.[8] We classified them in terms of neo-Kohlbergian stages, that are explained below in the discussion. Three such types can be distinguished (see Figure 2):

(1) those who ignore or fail to see the conflict of interest inherent in the PD and just do what they think is best for all, i.e., to cooperate (Stage 1C).
(2) those who see the conflict of interest and argue that they just have to pursue theirs, which is to defect (Stage 2A).
(3) those who see the conflict of interest and wish to overcome it. They decide to cooperate, but also express that they would be ready to defect, should their partner be unwilling to cooperate (Stage 2C).

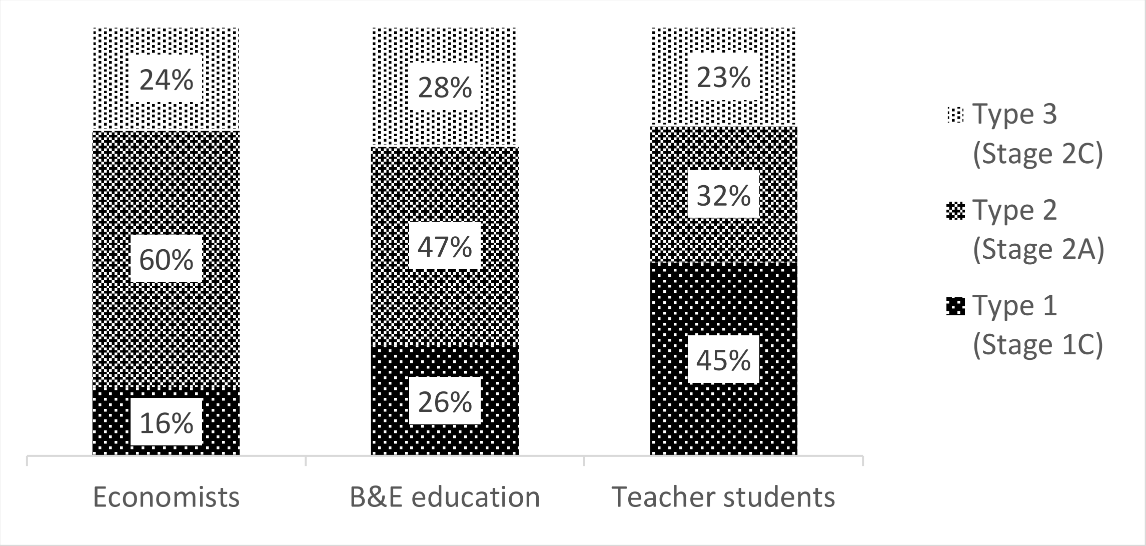

Figure 2: Types of Moral Reasoning.

[8] Since both kinds of reasoning complemented each other, we have taken them together as expressing their moral judgement.

With respect to the third type, there is no difference between the groups. Therefore, the overall difference in moral orientations is down to a trade-off between the other two types. Economists have a comparatively strong tendency to argue that as an agent in this game they have to pursue their personal interest, whereas teacher students show a strong orientation towards what would be collectively rational (χ^2 = 29.548, *df* = 4, *n* = 338, *p* = 0.000). If we control for gender, however, we only get significant results for female participants (χ^2 = 17.653, *df* = 4, *n* = 205, *p* < 0.001; male: χ^2 = 7.648, *df* = 4, *n* = 130, *p* = 0.053. one-tailed). Among economists, 60.2 percent are type 2, and 16.1 percent are type 1. Conversely, among teacher students we find 32.2 percent type 2 and 44.8 percent type 1 participants.

The fact that there are no differences with respect to the more sophisticated third type is at least an indication that the participants might not differ in their moral preferences, but instead in their relevant beliefs. The strong defective orientation, coupled with type-2-reasoning that economists reveal, may be due to the fact that they are highly aware of the situational constraint the PD poses. Conversely, the teacher students seem to be oblivious to this very fact. They seem to focus on the common interest, ignoring that the PD models conflicting interests. In other words, the common interest should be to overcome this problem of conflicting interests, which is precisely what type-3-participants aim at. However, type-1-participants are unable (or unwilling) to see this basic problem. In the context of the differentiation between preferences and restrictions this means that they, as a matter of fact, conceive of quite different restrictions than participants of type 2 or type 3.

Inasmuch as the differences in agency derive from differences in beliefs rather than differences in basic moral preferences, the different groups might differ neither in their moral judgement competence nor in what might be called their moral motivation, but in their comprehension of situational constraints.

A further analysis, illustrated in Figure 3 supports this view. Here we only look at HVs and the reasons they give for their choices and feelings. Some of them explicitly refer to the situational constraints and state that cooperation would have been better, but in the situation as it was, they just had to choose the other option. They clearly would have preferred to cooperate. They are not unhappy, however, because they think they have taken the right decision. In quite some participants we could identify this kind of reasoning, and we call these "strategic moralists", because they are morally motivated, but also focus on the prudential aspect of what kind of morality can be implemented in the present situation.

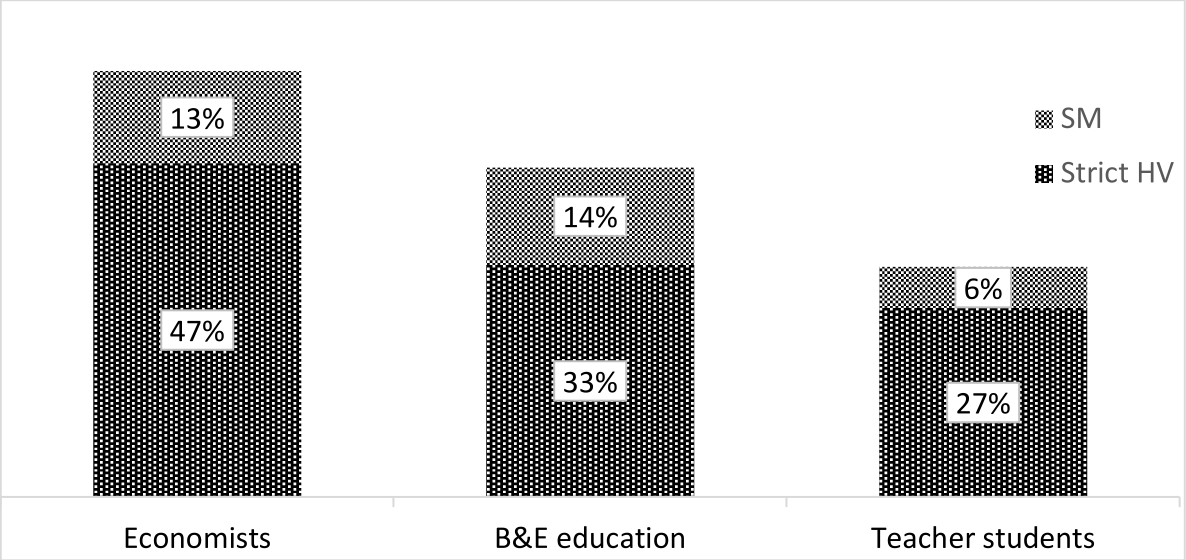

Figure 3: Strategic Moralists and Happy Victimizers in the Strict Sense.

Again, we see more strategic moralists among economists than among teacher students. This does not mean that economists are, generally speaking, not more selfish than others. Hypothesis 3 only states that the stronger tendency towards an HV-like agency on the part of the economists is not only an effect of a higher level of selfishness, but in part down to different beliefs. This is confirmed by the data. Furthermore, students of B&E education score on intermediate levels also in this respect. Hypothesis 4 is thus confirmed throughout.

4. Discussion: Moral Principles as Social Institutions

4.1. Morality in the Prisoners' Dilemma

Above, I have made a claim for a game-theoretic understanding of morality, where moral rules have to be rules of a game that the agents play. If they are to be rules in the game-theoretic sense, however, they have to be self-enforcing (Binmore, 2010). That is, following moral rules must be somehow pay – especially in "moral currencies" like respect, reputation and the like – and violating them must translate into costs, so that not complying with the rules is clearly irrational (at least for those who understand the rules).[9]

Put in other words, a simple appeal to moral rules and precincts is worth nothing, if those rules cannot secure compliance by themselves (and if this fault is not compensated by some other rules). This is the case in the (one-shot) PD, where a morality of contract or the Golden Rule cannot be implemented, simply because these moral rules are either impossible (striking a deal) or not enforceable (Golden Rule). The Golden Rule is not enforceable, because no signals of approval or disproval, or even of the appropriateness of the Golden Rule, can be exchanged, which means that this "moral currency" is invalid in this case.[10] This explains why such moral orientations are usually crowded out very quickly (for the remaining levels of cooperation see footnote 10).

However, this is not the whole story. If we take moral rules as rules of the game, we can make even more sense of the strategies that individuals choose in the game. *First of all*, even where individuals defect, they may still follow a moral rule, the rule that everyone has interests to pursue. This is one variant of Kohlberg's Stage 2, which concerns the "awareness that each person has interests to pursue and that these may conflict" (Colby & Kohlberg, 1987, p. 26). Defectors, whom we have classified as HVs, do not betray morality, but follow a specific morality that acknowledges the dignity of each person's individual interest and that these interests may conflict. This is the "type 2" morality specified above. In such cases of conflict of interest (and in the absence of any means to mediate between these conflicting interests), agents are morally justified to pursue their interests, but also have to accept others doing so. This morality is employed by both HVs and SMs, where the difference is that SMs explicitly state that a higher order morality would not work.

Secondly, "type 1" morality seems naïve, because it ignores the conflict of interest and rather assumes that each individual's interest in the other person's well-being is strong enough to preclude defection. This is a kind of morality that would work among closely affiliated agents (e.g., family members and friends are typically supposed to care for each other and dispense with a personal profit or benefit to support the other). Their moral appeals for solidarity would therefore fall flat.

[9] If one player does not understand the rules, the game is not played (at least not with this player), because a game implies that the players know the rules. If they don't, they may still play a game, but a different one.

[10] Of course, self-signalling is always possible, and this may be the mechanism that makes some people cooperate, even after having gained experience with the PD (Ledyard, 1995). Some take this as a strong moral self (Blasi, 1984; 1995; Bergman, 2002; 2004; Krettenauer, 2013). However, since this is tantamount to self-exploitation, one is certainly not morally obliged to cooperate in the face of (a high risk of) others defecting, especially since such defection cannot be penalised in any way. Hence, whether one should engage in one-tailes cooperation is a question of personal value and prudential reasoning, but not a moral issue in the strict sense, even though we often associate it with morality.

Thirdly, “type 3” morality is employed by those who see a chance to coordinate with others, but this higher form of morality is almost certain to be crowded out, since individuals frequently express they would switch to defection, if their partners fail to cooperate.

Fourthly, the situation would be different, if we allowed for changes in the game. For instance, the PD could be played repeatedly. In this case, a tit-for-tat strategy is possible. If player 2 understands the intention of player 1’s choices, cooperation can easily emerge in this repeated interaction. Player 1’s cooperation in the first round can be rewarded by player 2’s cooperation in the second round, and vice versa. Conversely, player 1’s defection in one round can be punished by player 2’s defection in the following round. Every cooperative choice can be understood as a promise to cooperate in the future, provided that the other player follows suit. Hence, under these circumstances the morality of promise-giving can be implemented und generate benefits for both

4.2. The Economics of Morality in the Game-theoretic Context

Finally, we can generalise the kind of moral functioning just explained with respect to the PD (see also Minnameier, 2018). The PD is one example of what game-theorists call a “cooperation game”. Perhaps counterintuitively, cooperation games are the situations in which cooperation typically fails if no institution is established, because the Pareto-efficient point is not a Nash-equilibrium. This formal structure of cooperation games is illustrated in Figure 4.

		Other	
		Defect	Cooperate
Self	Cooperate	1, 4 (lose-win)	3, 3 (win-win)
	Defect	2, 2 (lose-lose)	4, 1 (win-lose)

Figure 4: The Basic Structure of Cooperation Games.

The simplest version of such a game is the situation that Thomas Hobbes models in the *Leviathan*: Based on the “law of Nature” according to which “every man has right to everything” (1651/2001, p. 65 [Chap. 15, §2]) a “war of every one against every one“ (1651/2001, p. 59 [Chap. 14, §4]) is thought to ensue. This marks the very beginning of morality, where, e.g., children get into conflict about some resources like food or toys and they compete in trying to appropriate things.

What they have to learn is to mutually respect property or rights of use (e.g. that the one who had it first has the right to use a certain item). However, before they learn this, they have to experience the social trap (i.e., the inefficient Nash equilibrium) they reach in the *bellum omnium contra omnes*, for this is the problem that calls for innovation. At the same time, this problem allows them to take the perspective of the other individual who has the opposite point of view. Both “players” understand that when they win something, the other one loses it (win-lose), and vice versa (lose-win). And eventually they learn that their conjoint activity produces a lose-lose-result. Establishing the moral norm allows them to move into the win-win-zone.

Since it applies to a cooperation game, a moral norm – as an “institution” – has to go with the possibility of sanctions. In the simplest case of morality in narrow social relationships, the agents sympathise with each other, want to make and keep friends and to win each other’s affections. Therefore, signs of fondness and attachment, like smiles and so on, function as positive sanctions, whereas repudiation and anger function as negative sanctions. If the sanctions work this way in a social relationship, the payoff matrix is changed accordingly (see Figure 5).

		Other	
		Defect	Cooperate
Self	Cooperate	1+3, 4-3 (win-lose)	3, 3 (win-win)
	Defect	2, 2 (lose-lose)	4-3, 1+3 (lose-win)

Figure 5: Payoff Matrix with Premiums (+3) and Discounts (-3) for Sanctions.

If the institution, i.e., the moral rule, is understood by both players, defection is discounted by the costs for spoiling the personal relationship and incurring dislike. Conversely, cooperation is enhanced by the returns in affection currency.

The most important aspect, however, of this change is that the whole game changes decisively. It is no longer a cooperation game, but a so-called coordination game in which the Pareto-efficient combination is also a Nash-equilibrium. The main result of this analysis is, therefore, that moral rules allow us to turn *cooperation games* into *coordination games*, and in this very sense they become self-enforcing, as long as the players understand them and as long as the sanctions work. The latter explains, why morality frequently vanishes in situations characterised by anonymity and social distance (see e.g., Dana, Weber, & Kuang, 2007; Andreoni & Bernheim, 2009).

4.3 Moral stages and appropriate sanctions

The economic rationale just sketched allows us to construct ever more complex kinds of social interaction and cooperation. This entails a succession of cooperation games that have to be turned into coordination games with the help of moral principles as the regulating institutions. Another property of this succession of stages is their dialectic order. If mutual acceptance of property rights marks the beginning of moral reasoning, this means that the perspectives and legitimate claims of different individuals are treated equally, each in their own right.

However, these perspectives are kept independent of each other, and the morality of property rights secures and underpins this independence. This principle ceases to fit, if individuals differ in their property rights. If one individual owns a certain item desired by the other(s), a sharing norm is what we need. Therefore, we usually claim that one ought to share, at least with friends or with those who are dear to us.

The sharing norm relates these perspective reciprocally to each other. However, there is a simple form of sharing, where those who have something have to give to others and have to share in equal parts, usually. And there is an advanced form of sharing, where relevant inter-individual differences (of deservingness in terms of need, effort, and so on) have to be taken into account. In this latter case, equal sharing is considered unjust, and a norm of equitable sharing has to be put in place.

This basic triad of types of morality follows from a dialectical approach advocated by the mature Piaget in one of his last works (Piaget & Garcia, 1989), where he calls them “intra”, “inter”, and “trans”, respectively. Empirical evidence is available today, e.g., from experiments with young children between three and six years of age (Paulus & Moore, 2014). Another tenet of Piaget’s final version of the mind’s architecture is that the "trans"-stage forms a complex unity which however can break up into a set of trans-stages as a further morally relevant aspect enters the scene that cannot be integrated into the present trans-stage. And so a higher level with a new stage triad opens up.

In this very sense, the early morality, which is based on sympathy and altruism, gives way to new and higher form of morality relating to the cases where helping someone else is futile and goes against one's own interest. In early forms of morality, helping others is never in contradiction with one's own interest, i.e., one enjoys helping others and caring for them. However, if you ought to care for a competitor in business or in sport, or where someone, say, wants you to go for a walk together, when you yourself prefer to stay at home, this marks a true clash of interests.

According to Kohlberg, the morality of conflicting interests is captured by his Stage 2 (Colby & Kohlberg, 1987). And since Kohlberg had also invented the idea of sub-stages "A" and "B" for every stage – that he later changed into "types"[11] – I label the first triad as Stages 1A, 1B, and 1C, and the succeeding triad Stage 2A, 2B, and 2C, and so on. In this so-called neo-Kohlbergian framework we find 9 stages altogether (see Minnameier, 2014). They cannot be explained in detail here, but the first three of them are briefly described in Table 2, together with the sanctions that work in its respective context.[12]

Table 2

: Moral Principles and Sanctioning Potentials for Stages 1 to 3

Stage	Principle	Reward	Punishment	Ext'd punishment (one stage down)
1A	Property	Affection	Dislike	Revenge (→ 0)
1B	Sharing/turn taking	Affection	Dislike	Stop sharing (→ 1A)
1C	Care/equity	Affection	Dislike	Strict reciprocity (→ 1B)
2A	Legitimate Interest	Respect	Disrespect	Suspension/separation (→ 1C)
2B	Promise/contract	Respect	Disrespect	Defection in PD/self interest (→ 2A)
2C	Golden rule	Respect	Disrespect	Tit-for-tat in PD (→ 2B)
3A	Group norms/roles	Repute	Disrepute	Distrust in group/exclusion (→ 2C)
3B	Cultural norms/expect.	Repute	Disrepute	Distrust in public/retaliation (→ 3A)
3C	Legal justice	Repute	Disrepute	Decent legal punishment (→ 3B)

Since Stage 1 is based on sympathy, sanctions are imposed in terms of affection (positive) and dislike (negative). This is the code in which moral discourse takes place at this stage (or the "moral currency" of Stage 1). If moral discourse in terms of this stage – or any other stage – fails because of unwillingness or inability, one can always back out of "the game", shift to a lower stage and thus play a different game. For instance, if others simply wouldn't stop taking things away from me or using and possibly spoiling my property illicitly, I will have to fight back in some way.

[11] Kohlberg first introduced these forms as „sub-stages" (see e.g. 1984), but later treated them as mere „types", because he noticed anomalies in the developmental sequence (Colby & Kohlberg, 1987). From the point of view of the neo-Kohlbergian taxonomy, however, these anomalies are integrated and therefore constitute no systematic problem anymore.

[12] Neo-Kohlbergian stages roughly conform to Kohlberg's original stages (at least with respect to stages 1 through 5). However, there are a few important differences with respect to particular substages. For instance, the "golden rule" integrates conflicts of interests and identified as Stage 2C in the Neo-Kohlbergian framework, while Kohlberg takes it as a form of Stage 3.

At Stage 1 this always relates to the people with whom one wants or has to keep up (like parents, mates and so on). Thus, the extended punishment at Stage 1A means to shift to Stage 0, i.e., to defend oneself by taking vengeance (this is the Hobbesian state of nature). Similarly, at Stage 1B one can ultimately stop sharing, so that everybody sticks with what they have (which is the principle of Stage 1A), and at Stage 1C one can always revert to strict reciprocity rather than an equitable or caring interaction.

At Stage 2 the legitimacy of interests and the conflicts of interest that arise are the core problem. This stage applies to situations in which the agents are mutually disinterested, either with respect to the person of the other or with respect to a specific activity. Thus, while siblings are generally quite interested in each other and are involved in a close relationship, they may none the less have diverging interests in terms of leisure time activities, and then it is legitimate for them to go their own sweet ways, as it were. Accordingly, a conflict may arise, when one, e.g., wants to listen to loud music while the other has to prepare for an exam (the same of course applies to students sharing a flat). If no ways can be found in which each one can pursue their interests without encroaching upon the others', one has to go separate ways. This is meant by "suspension" or "separation" as the extended punishment at Stage 2A, where the agents are relegated to contexts, where they do not interfere anymore with each other and only deal with those with whom one gets on well and wants to affiliate.

Stages 2B and 2C provide forms of coordinating in conflicts of interest. Mutual promises at Stage 2B agents with different interests to strike deals so that everybody's interest is furthered. Any ordinary deal is an example. The PD is a situation in which the agents would like to strike such a deal to overcome the dilemma. But the rules of this game preclude this. Therefore, agents have to revert to Stage 2A and simply pursue their own interest (by defecting). In real life this can be a form of punishing those who go back on their promises. In the PD it is morally just to act selfishly, because the conflict of interest cannot be resolved. Stage 2B requires that both parties have something to trade. Stage 2C goes beyond this, because here one makes a contract with oneself, determining what you would have yourself do, if you were the other person. However, this requires trustful relationships and that the favours you offer are paid back in case the roles were reversed some other time when you are in need of assistance or so. In case of violation of this principle one could revert to 2B and demand immediate compensation for any service to be given.[13]

At Stage 3A individual interests are merged into group interests. In a way, Stage 2C implies that one tries to please everybody and coordinate diverging individual interests in this way. However, there are situations in which this is impossible, e.g., if one works for a company where pleasing customers of suppliers too much means to spoil the company's business. At this point it is important to think in terms of social units like companies, departments, families and clans, or peer groups and teams (in sport or at work). Stage 3A relates to these social units and the roles one takes on in these contexts. Role-related reputation is what one can gain, and disrepute may be the price one has to pay, if one fails to fulfil one's tasks. In the limiting case, one is excluded from the group – either literally, e.g., by being laid off from a firm, or in the sense that group cohesion breaks apart. In the latter case, one can still treat each other with respect and rely on each other in terms of Stage 2C, but not in terms of role-related duties and commitments.

Stage 3B concerns inter-group relationships, with customs and generalised expectations as moral principles. Examples are fairness rules in sport, social "conventions"[14] like what kinds of behaviour are

[13] Kohlberg associated the Golden Rule with Stage 3. However, here he seems mistaken, since the Golden Rule applies to balance individual interests, whereas – at least in the neo-Kohlbergian framework – Stage 3 is based on *social units* to which individual interests are merged. In this sense any Stage 3 morality differs sharply from the Golden Rule, which, however, is integrated in this higher order of hierarchical complexity.

[14] Some separate strictly between moral issues and conventional issues (most prominently Turiel, 1983; 2002; Nucci, 2008). With respect to the systematic differentiation between "norms" and "conventions" in game theory (see above), I fully endorse this. However, social conventions can also function as norms in the sense and in the contexts discussed here. For further discussions of conventions functioning as norms see Bicchieri (2006) and Sugden (2010, where the latter explicitly discusses Turiel's and Nucci's approach).

expected from superiors and subordinates, honest practices in commercial relations (e.g., whether a handshake is an obligation or not).

Where there are diverging views among honest people – i.e., "honest" in the sense of Stage 3B – about what is decent and what not, an authority will have to decide, typically the leader or leading authority of group, which can be a political community, a company or any other kind of organisation. That there is a leader or an authority who knows (or must know) what is right and may legitimately decide, is the core of Stage 3C. At this stage, the verdict of an authority is believed to be sacrosanct.

At this point I end the description of moral stages. However, this is not to be mistaken as the endpoint. An authority as just described has to be legitimate, and what follows the acceptance of some kind of leadership is the question how we can determine whether certain laws, rules and forms of government are legitimate or not. This is the proper field of ethics that we would then enter. Altogether, the neo-Kohlbergian taxonomy of stages comprises 9 stages (Minnameier, 2000; 2001; 2005).

Finally, turning back to the three stages just described and to the way the sanctions work, I would like to point out that there are always two kinds of punishment available (see Table 2): One is punishment within an institution, which relates to the proper meaning of the moral principle and reclaims that it be heeded. If the other does not play according to these rules, one can still revert to a lower stage and consequently play another moral game at this lower level. This precludes self-exploitation in situations where certain moral rules cannot be implemented for some reason.

5. Conclusion

HVP does not seem to be a real problem. When people illegitimately act in selfish ways, well-functioning social systems should be able to counter such behaviour by appropriate sanctions. And if these sanctions fail to work effectively, we have to adapt our tools, in particular by playing different (lower-stage) moral games with other kinds of sanctions (as explained in section 5.3). Conversely, however, we will not solve such problems by trying to increase moral motivation in the sense criticised in this article (least of all among moral transgressors).

There are situations, where "happy victimizing" actually seems to be an appropriate behavioural orientation, in particular in strictly competitive situations, as when applying for a job or competing over a large order in business. In such situations, it is morally mandatory to pursue one's self-interest (as long as one is playing fair). The same is true for situations in which contracts are either not fulfilled or not feasible (like e. g., in the PD).

Keypoints

- The classical (morally externalistic) explanation based on moral motivation has been criticised and confronted with an internalistic alternative that incorporates a theory of inferential reasoning and a reason-based theory of rational choice.
- The reason-based approach and moral functioning can also be interpreted in terms of game theory, where moral principles (or preferences) become preferences for games.
- Empirical evidence supports the reason-based approach, which is then extended to an overall theory of moral principles as institutions in the institutional-economic sense.

- One main outcome of this analysis is that morality requires positive and negative sanctioning mechanisms that must be operative, if specific types of morality are to be upheld in specific contexts.
- This new approach also revives Kohlbergian moral theory and leads to a Neo-Kohlbergian theory of moral reasoning and moral functioning.

Acknowledgements

I wish to thank two anonymous reviewers who have read the paper very thoroughly and carefully. They have helped me to improve it and correct remaining errors.

References

Ameriks, K. (2006). Kant and motivational externalism. In H. F. Klemme, M Kühn & D. Schönecker (Eds.), *Moralische Motivation: Kant und die Alternativen* (pp. 3-22), Hamburg: Meiner.

Andreoni J., & Bernheim, D. B. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica, 77,* 1607–1636. https://doi.org/10.3982/ECTA7384

Ariely, D. (2012). The (honest) truth about dishonesty: How we lie to everyone – especially ourselves. New York: Harper Collins.

Arsenio, W. F., Gold, J., & Adams, E. (2006). Children's conceptions and displays of moral emotion. In: M. Killen/J. G. Smetana (Eds.). *Handbook of moral development* (pp. 581-609). Mahwah, NJ: Erlbaum.

Batson, C. D., Thomson, E. R., & Chen, H. (2002). Moral hypocrisy: Addressing some alternatives. *Journal of Personality and Social Psychology, 83,* 330-339. https://doi.org/10.1037/0022-3514.83.2.330

Batson, C. D., Thomson, E. R., Seuferling, G., Whitney, H, & Strongman, J. A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology, 77,* 525-537. https://doi.org/10.1037/0022-3514.77.3.525

Becker, G. S. (1976). *The economic approach to human behavior.* Chicago, IL: University of Chicago Press.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of Political Economy, 101*, 385–409.

Bergman, R. (2002). Why be moral? A conceptual model from developmental psychology. *Human Development, 45*, 104-124.

Bergman, R. (2004). Identity as motivation: Toward a theory of the moral self. In D. K. Lapsley & D. Narvaez (Eds.), *Moral development, self, and identity* (pp. 21-46), Mahwah, NJ: Lawrence Erlbaum Associates.

Bicchieri, C. (2006). *The grammar of society: The nature and dynamics of social norms*. Cambridge, UK: Cambridge University Press.

Binmore, K. (2009). *Rational decisions*. Princeton: Princeton University Press.

Binmore, K. (2010). Game theory and institutions. *Journal of Comparative Economics, 38,* 245-252. https://doi.org/10.1016/j.jce.2010.07.003

Blasi, A. (1984). Moral identity: Its role in moral functioning. In W. M. Kurtinez & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp. 128-139). New York: Wiley.

Blasi, A. (1995). Moral understanding and the moral personality: The process of moral integration. In W. M. Kurtinez & J. L. Gewirtz (Eds.), *Moral development: An introduction* (pp. 229-253). Boston: Allyn and Bacon.

Brickhouse, T. C., & Smith, N. D. (2010*). Socratic moral psychology.* Cambridge: Cambridge University Press.

Brink, D. O. (1997). Moral motivation. *Ethics*, *108*, 4-32.

Carter, J. R., & Irons, M. (1991). Are economists different, and if so, why? *Journal of Economic Perspectives, 5*(2)*,* 171–177.

Colby, A., & Kohlberg, L. (1987). *The measurement of moral judgment, Vol. I: Theoretical foundations and research validation.* Cambridge, MA: Cambridge University Press.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory, 33*, 67–80. doi: 10.1007/s00199-006-0153-z

Dietrich, F., & List, D. (2013a). A reason-based theory of rational choice. *Noûs, 47*, 104-134. https://doi.org/10.1111/j.1468-0068.2011.00840.x

Dietrich, F., & List, D. (2013b). Where do preferences come from? *International Journal of Game Theory, 42*, 613–637. doi: 10.1007/s00182-012-0333-y

Foot, P. (1972). Morality as a system of hypothetical imperatives. *Philosophical Review, 81*, 305-316. doi: 10.2307/2184328

Frank, R. H., Gilovich, T., & Regan, D. T. (1993). Does studying economics inhibit cooperation? *Journal of Economic Perspectives, 7*(2)*,* 159-171.

Frank, R. H., Gilovich, T., & Regan, D. T. (1996). Do economists make bad citizens? *Journal of Economic Perspectives, 10*(1)*,* 187-192. doi: 10.1257/jep.10.1.187

Frey, B. S. (1986). Economists favour the price system. Who else does? *Kyklos, 39(4),* 537–563. https://doi.org/10.1111/j.1467-6435.1986.tb00677.x

Frey, B. S., & Meier, S. (2003). Are political economists selfish and indoctrinated? Evidence from a natural experiment. *Economic Inquiry, 41,* 448-462. https://doi.org/10.1093/ei/cbg020

Gul, F., & Pesendorfer, W. (2008). The case for mindless economics. In A. Caplin & A. Schotter (Eds.), T*he foundations of positive and normative economics: A handbook* (pp. 3-39). Oxford: Oxford University Press.

Heinrichs, K., Minnameier, G., Gutzwiller-Helfenfinger, E., & Latzko, B. (2015). „Don't worry, be happy"? – Das Happy-Victimizer-Phänomen im berufs- und wirtschaftspädagogischen Kontext. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *111*, 31-55**.**

Hermkes, R. (2016). Perception, abduction, and tacit inference. In L. Magnani & C. Casadio (Eds.), *Model-based reasoning in science and technology – Logical, epistemological, and cognitive issues* (pp. 399-418). Heidelberg: Springer.

Hobbes, T. (1651/2001). *Leviathan*. South Bend, IN: Infomotions.

Kant, I. (2002/1785). *Groundwork for the metaphysics of morals* (ed. and transl. by A. W. Wood). New Haven, CT: Yale University Press.

Keller, M., Lourenço, O., Malti, T., & Saalbach, H. (2003). The multifaceted phenomenon of „happy victimizers": A cross-cultural comparison of moral emotions, *British Journal of Developmental Psychology*, 21, 1-18. doi: 10.1348/026151003321164582

Killen, M., & Smetana, J. G. (2015). Origins and development of morality. In M. E. Lamb (Ed.), *Handbook of child psychology and developmental science, Vol. 3* (7th ed.; pp. 701-749). NY: Wiley-Blackwell. https://doi.org/10.1002/9781118963418.childpsy317

Kohlberg, L. (1984). *Essays on moral development, Vol. 2: The psychology of moral development*. San Francisco, CA: Harper & Row.

Krebs, D. L., & Denton, K. (2005). Toward a more pragmatic approach to morality: A critical evaluation of Kohlberg's model, *Psychological Review*, *112*, 629-649. https://doi.org/10.1037/0033-295X.112.3.629

Krettenauer, T. (2013). Moral motivation, responsibility and the development of the moral self. In F. Oser, K. Heinrichs & T. Lovat (Eds.), *Handbook of moral motivation: Theories, models, applications*. (pp. 215-228). Rotterdam: Sense.

Krettenauer, T., Malti, T., & Sokol, B. W. (2008). The development of moral emotion expectancies and the happy victimizer phenomenon: A critical review of theory and application, *European Journal of Developmental Science, 2*, 221-235. doi: 10.3233/DEV-2008-2303

Ledyard, J. (1995). Public goods: A survey of experimental research. In J. Kagel & A. Roth (Eds.), *Handbook of experimental economics* (pp. 253–279). Princeton: Princeton University Press.

Malti, T., & Krettenauer, T. (2013). The relation of moral emotion attributions to prosocial and antisocial behavior: A meta-analysis. *Child Development, 84*, 397-412. doi: 10.1111/j.1467-8624.2012.01851.x

Marwell, G., & Ames, R. (1981). Economists free ride, does anyone else? *Journal of Public Economics, 15*(3)*,* 295–310.

Minnameier, G. (2000). *Strukturgenese moralischen Denkens - Eine Rekonstruktion der Piagetschen Entwicklungslogik und ihre moraltheoretischen Folgen*. Münster: Waxmann.

Minnameier, G. (2001). A new stairway to moral heaven – A systematic reconstruction of stages of moral thinking based on a Piagetian 'logic' of cognitive development. *Journal of Moral Education*, *30*, 317-337. https://doi.org/10.1080/03057240120094823

Minnameier, G. (2005). Developmental progress in Ancient Greek ethics. *European Journal of Developmental Psychology*, *2*, 71-99. https://doi.org/10.1080/17405620444000274a

Minnameier, G. (2010). The problem of moral motivation and the Happy Victimizer Phenomenon – Killing two birds with one stone. *New Directions for Child and Adolescent Development, 129*, 55-75. https://doi.org/10.1002/cd.275

Minnameier, G. (2012). A cognitive approach to the 'happy victimiser'. *Journal of Moral Education*, *41*, 491-508. https://doi.org/10.1080/03057240.2012.700893

Minnameier, G. (2013). Deontic and responsibility judgments: An inferential analysis. In F. Oser, K. Heinrichs & T. Lovat (Eds.), *Handbook of moral motivation: Theories, models, applications*. (pp. 69-82). Rotterdam: Sense.

Minnameier, G. (2014). Moral aspects of professions and professional practice. In S. Billet, C. Harteis & H. Gruber (Eds.), *International handbook of research in professional and practice-based learning* (pp. 57-77). Berlin: Springer.

Minnameier, G. (2016a). Rationalität und Moralität – Zum systematischen Ort der Moral im Kontext von Präferenzen und Restriktionen. *Zeitschrift für Wirtschafts- und Unternehmensethik, 17*, 259-285.

Minnameier, G. (2016b). Abduction, selection, and selective abduction. In L. Magnani & C. Casadio (Eds.), *Model-based reasoning in science and technology – Logical, epistemological, and cognitive issues* (pp. 309-318). Heidelberg: Springer.

Minnameier, G. (2017). Forms of abduction and an inferential taxonomy. In L. Magnani & T. Bertolotti (Eds.), *Springer Handbook of Model-Based Reasoning* (pp. 175-195). Berlin: Springer.

Minnameier, G. (2018). Reconciling morality and rationality – Positive learning in the moral domain. In O. Zlatkin-Troitschanskaia, G. Wittum & A. Dengel (Eds.), *Positive learning in the age of information (PLATO) - A blessing or a curse?* (pp. 347-361). Wiesbaden: Springer VS.

Minnameier, G., & Schmidt, S. (2013). Situational moral adjustment and the happy victimizer. *European Journal of Developmental Psychology, 10*, 253-268. doi: 10.1080/17405629.2013.765797

Minnameier, G., Beck, K., Heinrichs, K., & Parche-Kawik, K. (1999). Homogeneity of moral judgement? Apprentices solving business conflicts. *Journal of Moral Education*, *28*, 429-443. https://doi.org/10.1080/030572499102990

Minnameier, G., Heinrichs, K., & Kirschbaum, F. (2016). Sozialkompetenz als Moralkompetenz – Theoretische und empirische Analysen. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 112*, 636-666.

Nucci, L. (2008). Social cognitive domain theory and moral education. In L. Nucci, & D. Narvaez (Eds.), *Handbook of moral development and character education* (pp. 291–309). Oxford: Routledge.

Nunner-Winkler, G. (1999). Development of moral understanding and moral motivation. In F. E. Weinert & W. Schneider (Eds.), *Individual development from 3 to 12* (pp. 253–292). Cambridge, UK: Cambridge University Press.

Nunner-Winkler, G. (2007). Development of moral motivation from childhood to early adulthood. *Journal of Moral Education, 36*, 399-414. https://doi.org/10.1080/03057240701687970

Nunner-Winkler, G. (2013). Moral motivation and the happy victimizer phenomenon. In F. Oser, K. Heinrichs & T. Lovat (Eds.), *Handbook of moral motivation: Theories, models, applications.* (pp. 267-288). Rotterdam: Sense.

Nunner-Winkler, G., & Sodian, B. (1988). Children's understanding of moral emotions, *Child Development, 59*, 1323-1338. doi: 10.2307/1130495

Paulus, M. (2014). The emergence of prosocial behavior: Why do infants and toddlers help, comfort, and share? *Child Development Perspectives, 8,* 77-81. https://doi.org/10.1111/cdep.12066

Paulus, M., & Moore, C. (2014). The development of recipient-dependent sharing behaviour and sharing expectations in preschool children. *Developmental Psychology, 50*, 914-921. https://doi.org/10.1037/a0034169

Piaget, J., & Garcia, R. (1989). *Psychogenesis and the history of science.* New York: Columbia University Press.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review, 118,* 57-75. https://doi.org/10.1037/a0021867

Rest, J. R. (1984). The major components of morality. In W. M. Kurtinez & J. L. Gewirtz (Eds.), *Morality, moral behavior, and moral development* (pp. 24-38). New York: Wiley.

Rosati, C. S. (2016). Moral motivation. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/win2016/entries/moral-motivation/>.

Rubinstein, A. (2006). A sceptic's comment on the study of economics. *Economic Journal, 116*, C1-C9. https://doi.org/10.1111/j.1468-0297.2006.01071.x

Rustichini, A., & Villeval, M. C. (2014). Moral hypocrisy, power, and social preferences. *Journal of Economic Behavior & Organization, 107*, 10-24. https://doi.org/10.1016/j.jebo.2014.08.002

Samuelson, P. A. (1938). A note on the pure theory of consumer's behaviour. *Economica, 5,* 61–71. doi: 10.2307/2548836

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica, 15,* 243–253. 10.2307/2549561

Selten, R., & Ockenfels, A. (1998). An experimental solidarity game. *Journal of Economic Behavior & Organization 34* (4), 517-539. https://doi.org/10.1016/S0167-2681(97)00107-8

Smith, M. (1994/2005). *The moral problem*. Oxford: Blackwell.

Sugden, R. (2010). Is there a distinction between morality and convention? In M. Baurmann, G. Brennan, R. E. Goodin & N. Southwood (Eds.), *Norms and values: The role of social norms as instruments of value realization* (pp. 47-65). Baden-Baden: Nomos.

Thoma, S. J., & Bebeau, M. J. (2013). Moral motivation and the four component model. In F. Oser, K. Heinrichs & T. Lovat (Eds.), *Handbook of moral motivation: Theories, models, applications.* (pp. 49-68). Rotterdam: Sense.

Turiel, E. (1983). *The development of social knowledge: Morality and convention.* New York: Cambridge University Press.

Turiel, E. (2002). *The culture of morality: Social development, context, and conflict.* New York: Cambridge Universtiy Press.

Warneken, F., & Tomasello, M. (2009). The roots of human altruism. *British Journal of Psychology*, *100*, 455–471. https://doi.org/10.1348/000712608X379061

Zangwill, N. (2003). Externalist moral motivation. *American Philosophical Quarterly, 40*, 143-154.