Competing Image Concerns: Pleasures of Skill and Moral Values

Armin Falk* and Nora Szech†

October 15, 2020

Abstract

We study how two competing sources of positive image – the desire to appear skillful vs. moral – affect moral decision making. In four experimental conditions we vary the perceived importance of skill as well as moral relevance. In all conditions, participants completed an IQ test. To vary perceived importance of the skill motive, we either framed the IQ test as an "IQ test", or as a simple "questionnaire", respectively. Variation in moral consequences was implemented in adding negative externalities in response to correctly answered questions. Our results show that adding moral consequences decreases the number of correctly answered questions, relative to the morally neutral conditions, revealing that participants care for a positive moral image. Importantly, however, moral outcomes are significantly less likely when the test is framed as an IQ-test relative to the questionnaire framing. Our findings show that the pursuit of pleasures of skill can causally reduce moral behavior.

JEL classification: D64, C91

Keywords: Moral behavior, Signaling, Pleasures of skill, Self-relevance, Experiments

*Institute of Behavior and Inequality (briq) and Department of Economics, University of Bonn.

†Department of Economic, Karlsruhe Institute of Technology, Berlin Social Science Center and CESifo.

We thank Benjamin Enke, Anke Greif-Winzrieth, and Jana Willrodt, and in particular, Luca Henkel, for excellent research assistance. We thank Douglas Bernheim, Vincent Crawford, Thomas Dohmen, Johannes Hermle, Paul Heidhues, David Huffman, Michael Kirchler and Matthias Stefan as well as audiences at University of Cologne, CREED, DICE, DIW Berlin, EEA-ESEM Geneva, SITE Psychology and Economics, HeiKaMax, and VJE Seminar at University of Vienna for helpful discussions. For technical, programming, or administrative support, we thank Markus Antony, Holger Gerhardt, and Sibille Hackel. Armin Falk acknowledges financial support by the German Science Foundation (Deutsche Forschungsgemeinschaft) through the Leibniz Program. The mouse paradigm was approved by the Ethics Committee of the University of Bonn (reference no. 066/12).

1 Introduction

"When you see something that is technically sweet, you go ahead and do it."

J. R. Oppenheimer, scientific leader of the Manhattan Project

This paper investigates how the pursuit of excellence can negatively affect the likelihood of moral decision making. We hypothesized that this might be the case when two fundamental motives of behavior, the desire to appear skillful and the desire to appear moral, are in conflict and call for opposing behaviors. A striking example from history which illustrates this tension is the development of the atomic bomb. The Manhattan project was one of the most challenging engineering, management and scientific projects of the twentieth century, bringing together leading physicists of the time and providing them with a unique opportunity to excel[1]. Physicist Richard Feynman stated about the Manhattan Project: "(...) we started for a good reason, then you're working very hard to accomplish something and it's a pleasure, it's excitement. And you stop thinking, you know; you just stop." (Feynman, 1985, p. 136). Even after knowing about the devastating power of the bomb, for "many of the physicists, the initial reaction was excitement over a spectacular demonstration of the successful 'technologically sweet' joint effort leading up to 'the gadget', as it was known" (Mårtensson–Pendrill, 2006).[2] However, after the dropping of the plutonium bomb on Nagasaki, which was deemed unjustified by many of the scientists, numerous members of the Manhattan Project started worrying about moral implications. Many suffered from feelings of depression, nausea, guilt and, in some cases, outright horror (Lifton and Mitchell, 1995; Russell, 1960; Weaver, Snow, Hesburg, and Baker, 1961).

The fact that humans like achievements as well as mastering and excelling in tasks has forcefully been pointed out by Bentham (1781), who, in his famous list of *14 Pleasures*, calls this desire the *pleasures of skill*.[3] In psychology, this motive has been identified as a human need for "effectance" described as the ability to successfully produce desired outcomes (Ahuvia, 2005; Belk, 1988; Dittmar

[1] Mårtensson–Pendrill (2006) describes Los Alamos, where the project took place as "(...) an exciting brewing pot of creative brains focusing on very difficult problems, involving both theoretical challenges, technological development, logistics and numerics."

[2] Another member of the Manhattan Project, Australian physicist Sir Mark Oliphant, commented that he "learned during the war that if (...) the work's exciting" scientists will "work on anything", and continued that there is "no difficulty getting doctors to work on chemical warfare and physicists to work on nuclear warfare" (Lifton and Mitchell, 1995).

[3] In a similar vein, Marshall (2009) argued that the "importance of the mastery motive has been a large part of the demand for the highly skilled professional services and the best work of the mechanical artisan arises from the delight people have in the training of their faculties and in exercising them."

and Pepper, 1992; Ewers and Zimmermann, 2015; Furby, 1991; Norton, Mochon, and Ariely, 2012; White, 1959). However, individuals not only strive for an image of being skillful; they also seek to pursue a positive moral image. Various papers have documented that individuals are eager to signal their prosociality either vis-a-vis others (e.g., Ariely, Bracha, and Meier, 2009; Butler, Serra, and Spagnolo, 2020; DellaVigna, List, and Malmendier, 2012; Gerber, Green, and Larimer, 2008; Lacetera, Macis, and Slonim, 2012), or themselves (e.g., Falk, 2018; Mazar, Amir, and Ariely, 2008). Common to the desire to signal either skillfulness or prosocial inclinations is the desire for maintaining a positive self-image (Bénabou and Tirole, 2005, 2006; Bénabou, Falk, and Tirole, 2020; Kuhnen and Tymula, 2012; Loewenstein, 1999; Prelec and Herrnstein, 1991). This desire may stem from an uncertainty about who we are, how skillful and how moral, but it is also instrumental, e.g., for getting better matches in job or partner markets. A potential threat to moral outcomes and welfare arises if individuals are called upon acting on a task that is simultaneously morally problematic, but also considered suitable for proving their skillfulness. To study an environment with competing image concerns, we implemented a choice context that varies both the relevance of signaling skills and being moral.

We study four experimental conditions (*2x2* design) and vary the perceived importance of skill as well as moral relevance. In all conditions, participants completed an IQ test. To vary perceived importance of the skill motive, we either framed the IQ test as an "IQ test", or as a simple "questionnaire", respectively. Given the importance of being intelligent, performing well on an IQ-test has a higher self-relevance compared to doing well on a simple questionnaire. With respect to variation in the moral domain, in two treatments, correctly answering questions in the test produced a negative externality. In particular, in these treatments participants knew that each correctly solved question increased the probability that a mouse is getting killed. With these variations we have four treatments, characterized by two types of framings (IQ vs. questionnaire) and with or without moral consequences. This allows us to draw inferences on whether pleasures of skill can causally undermine moral behavior. Our results first show that adding moral consequences decreases the number of correctly answered questions, relative to the two morally neutral conditions. This shows that participants care for a positive moral image. Importantly, however, the likelihood of moral outcomes is significantly lower when the test is framed as an IQ-test relative to the questionnaire framing. In particular, the reduction in correctly solved questions in response to facing a negative externality is significantly lower when participants make decisions in the IQ relative to the questionnaire conditions.

Our results show that moral outcomes are at risk in situations where the desire to appear moral competes with pleasures of skill. In this sense, they add to our understanding of the context specificity of moral behavior and how easily it is malleable. After all, the main treatment variation merely changes the framing and even if participants are eager to learn their IQ score, they could easily perform respective tests after participating in the experiment, creating no externalities. Our results also point to the importance of accounting for multiple sources of image. Previous research has documented the relevance of image concerns in single domain contexts, but decisions in many important life contexts allow inferences in multiple dimensions. Which dimension prevails in determining behavior is an empirical question, but it is important to keep potentially conflicting signals in mind when applying image concern models to complex contexts such as the workplace, the mating market or educational contexts.

The remainder of the paper is as follows: Section 2 provides a description of the design. Our results are presented in section 3, and section 4 concludes.

2 Experimental Design

To identify potentially adverse effects of pleasures of skill on moral behavior, in our two main treatments we implemented a morally problematic task and varied whether or not subjects consider the task self-relevant in terms of skillfulness (*IQ* and *NoIQ*). In order to confirm that the decision environment in our main treatments is of moral relevance to subjects, we ran two otherwise identical control conditions in a morally neutral environment (*Control_IQ* and *Control_NoIQ*). This constitutes a *2x2* experimental design, summarized in Table 1.

In all four conditions, subjects were administered the same task, a 52-item intelligence test. The test included items for fluid intelligence (Raven matrices) as well as crystallized intelligence (vocabulary test questions). Importantly, in all conditions subjects only received a show-up fee of 15 Euros, i.e., payment did not depend on test performance. Before starting the test, subjects were told that at the end of the test, they will be informed about their test score and whether their score was above or below the average performance of a large group of students who have previously

		Morality	
		Main treatments: Morally	**Control treatments: Morally neutral**
Pleasures of Skill	**High**	Treatment: *IQ* Framing: IQ test Morality: Killing of mice N = 84	Treatment: *Control_IQ* Framing: IQ test Morally neutral N = 72
	Low	Treatment: *NoIQ* Framing: Questionnaire Morality: Killing of mice N = 81	Treatment: *Control_NoIQ* Framing: Questionnaire Morally neutral N = 64

Table 1: Design overview and treatment variations

performed the same test in unrelated experiments.[4]

To vary the self-relevance of the test we framed it either as an "IQ test" or simply as a "questionnaire". In the IQ treatments *IQ* and *Control_IQ*, the instructions informed subjects that they are taking part in an IQ test, and that IQ is a crucial skill that is associated with important life outcomes such as wages or educational achievement. In contrast, in the treatments *NoIQ* and *Control_NoIQ* subjects were told that they are taking part in a questionnaire study, without alluding to intelligence. Hence, the difference between the IQ and NoIQ conditions was simply a matter of how the test was framed.

We chose an IQ test context for two reasons. First, IQ is arguably one of the most important personal characteristics in determining overall success in life (Borghans, Duckworth, Heckman, & Ter Weel, 2008; Sternberg, Grigorenko, & Bundy, 2001).[5] It has been found to reliably predict important life outcomes, such as educational achievement, employment prospects, wealth acquisition or career prospects. Therefore, individuals like to think of themselves as being intelligent. In this sense, the IQ context ensures high ecological validity in representing a measure of positive self-image with regard to skills. Second, using an IQ test allows us to easily vary the meaning and

[4]This average is actually based on test scores of about 1,700 students who have done the same test.

[5]IQ refers to the "ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought" (Official taskforce of the American Psychological Association; Neisser et al., 1996).

self-relevance of the task, by means of different framing, while at the same time exposing subjects to the identical task. The fact that treatment comparisons are not confounded by using different tasks across treatments is an essential feature of our experiment. Implementing different tasks, e.g., an IQ test versus a pointless and non-self-relevant task, would pose the problem of dealing with different metrics in the dependent variable and different production functions across tasks. This could bias findings in unknown ways, unrelated to our hypothesis. Using the same task in all conditions, the dependent variable and effort cost functions for answering the test questions are kept identical.

In the treatments *IQ* and *NoIQ*, moral consequences were implemented using the Mouse Paradigm (Falk & Szech, 2013). The paradigm involves the killing of mice, which is morally objectionable, since killing an animal implies causing harm in an intentional and unjustified way. While there exists no universal consensus about how to define the content of morality, avoiding and preventing harm is a central element according to most notions of morality (Gert & Gert, 2017).[6]

In the instructions, each subject was told that he or she is "endowed" with a mouse, whose life will depend on performance in the test. In particular, subjects were informed that for each correctly answered question in the test, the likelihood that their mouse will be killed increases by 0.9 percentage points. For example, correctly solving 10 questions implies a killing probability of 9 percent, solving 30 questions one of 27 percent and so one. Hence the maximum killing probability is 46.8 percent, in case all questions are answered correctly and it is zero if no question is solved correctly, respectively. To allow subjects to save their mouse with certainty, each of the 52 test items included the possible response option "don't know" (this holds for all four treatment conditions). This allowed subjects to answer each question incorrectly and to guarantee survival of their mouse, independently of whether they knew the correct answers to the test questions.

Subjects were informed that their mouse is a young and healthy mouse, which will live in an appropriate, enriched environment, jointly with a few other mice in case it is not killed. To rule out any uncertainty about consequences, the instructions also informed subjects explicitly about the

[6] In Deckers, Falk, Kosse, and Szech (2016), we provide convergent and discriminatory validity of the mouse paradigm as a measure for morality. Killing is negatively related to agreeableness, one of the Big-Five facets, which describes a tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others; and positively related with Machiavellianism, measuring a person's tendency to be unemotional, and detached from conventional morality. Moreover, killing is not related to disposable income, whether students are professionally involved with animal research or animal experiments or a simple preference for animals, as expressed by having a pet at home.

killing process: “The mouse is gassed. The gas flows slowly into the hermetically sealed cage. The gas leads to breathing arrest. At the point at which the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed.” The instructions also presented a picture of a mouse and a short demonstration video of the killing process. In the video, four mice first move vividly in the cage, then they successively slow down. Eventually they die, with their hearts beating visibly heavy and slow.

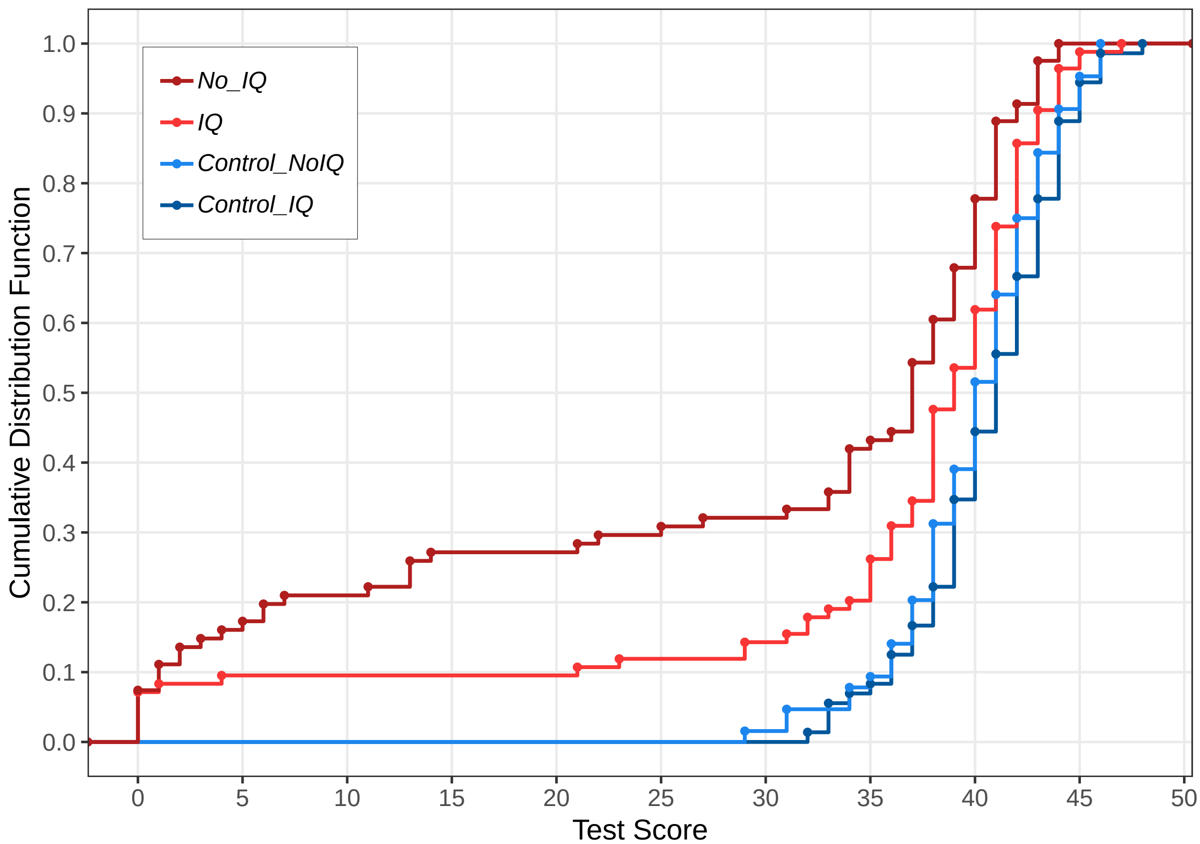
It is important to stress that all mice used in the experiment are so-called “surplus” mice. They were bred for animal experiments, but turned out unsuited or nor longer needed for particular studies. Such mice are perfectly healthy, but keeping them alive is costly. It is common practice in laboratories conducting animal experiments to gas surplus mice. Our study made it possible to rescue these animals. Surplus mice that were saved by subjects’ behavior were purchased by the experimenters and are kept in enriched conditions, precisely as stated in the instructions. Thus, as a consequence of our experiment, many mice that would have otherwise died were saved. Note that despite these mice being surplus mice, the decision of a study participant is about increasing the killing probability, i.e., a decision about life and death for a particular mouse.

A total of 301 subjects (123 male, mean age = 20.19, SD = 2.63) took part in the experiments, mainly undergraduate university students from various majors. All treatments were computerized using z-Tree as experimental software (Fischbacher, 2007) and subjects were recruited using the software ORSEE (Greiner, 2004). The experiments were conducted at the Bonn Econ Lab. A translation of the instructions is provided in the Appendix.

3 Results

Figure 1 displays the cumulative distribution functions of total test scores for all four treatment conditions. Several observations can be made. First, and as expected, adding a moral dimension reduces test scores. Comparing the two control treatments without moral consequences to those of our two main conditions with moral consequences shows that test scores are substantially lower in the latter. This difference is significant for both, the two IQ conditions ($p<0.01$, $n = 156$, two-sided rank sum test) as well as for the two NoIQ conditions ($p<0.01$, $n = 145$, two-sided rank sum test).

The important question, however, is whether the reduction in test scores in response to intro-



The figure displays empirical cumulative distribution functions of total test scores for the treatments with moral consequences (red) and without moral consequences (blue).

Figure 1: Pleasures of skill undermine moral behavior

ducing moral consequences is more pronounced in the IQ relative to the NoIQ conditions. Figure 1 as well as a comparison of means shows that this is the case. Test scores in *IQ* are 35.13 (SD: 12.09) on average, compared to 28.78 (SD: 15.62) in *Control_IQ*. Thus, relative to *Control_IQ*, subjects reach about 22 percent higher test scores. The difference between these two treatments is statistically significant ($p<0.01$, $n=165$, two-sided rank sum test). In Table 2 we report regression estimates to check whether the more pronounced reduction of test scores in *IQ* vs. *NoIQ* is statistically significant. We regress test scores on a *Mouse dummy*, an *IQ dummy* as well as on their interaction. The effect of adding moral consequences is large and significant, as indicated by the respective coefficient of the Mouse dummy. The IQ dummy is positive but not significant.[7] The

[7]This may seem surprising. Means and standard deviations of test scores in the control conditions are 40.7 (3.45) in *Control_IQ*, and 40.0 (3.64) in *Control_NoIQ*, respectively. Note, however, that subjects in the *NoIQ* condition were asked to take part in a questionnaire, without an interesting alternative and knowing that their performance would be compared to the average performance of others. This may have provided sufficiently strong incentives to

Dependent Variable:	Test Score
Mouse dummy	-11.222^{***}
	(1.795)
IQ dummy	0.667
	(0.611)
Mouse-IQ-Interaction	5.687^{**}
	(2.266)
Constant	40.0^{***}
	(0.454)
Observations	301
R-squared	0.171

Table 2: OLS regression coefficient estimates

Estimates with test score as dependent variable and robust standard errors in parentheses. *Mouse dummy* is a dummy, which takes value 1 if observations come from mouse treatments. *IQ dummy* is a dummy, which takes value 1 if observations come from IQ treatments. *Mouse-IQ-Interaction* is an interaction of *Mouse dummy* and *IQ dummy*. *** indicate significance at the 1-percent-level, ** indicate significance at the 5-percent-level.

coefficient of interest is the interaction term, indicating the difference-in-differences when adding a moral consequence in *IQ* vs. *NoIQ*. The respective coefficient is positive and statistically significant (p=0.013). Hence, immoral behavior is more likely in a self-relevant context that allows for signaling skills, in comparison to a choice context that limits the possible experience of pleasures of skill.

A closer inspection of test score distributions shown in Figure 1, indicates that no participant scored below 29 in the morally neutral conditions. This suggests that a test score of 29 can be considered as the minimum performance level subjects are cognitively able to achieve. A lower test score in the conditions with moral consequences can therefore be interpreted as reflecting a deliberate choice in favor of saving the mouse. In fact, in the two main conditions, a sizable fraction of subjects did score below this minimum level. In *NoIQ*, 32.1 percent of test scores are strictly below 29, significantly more than in *Control_NoIQ* (p<0.01, Two Sample Test of Proportion, two-sided, n=145). However, in *IQ*, only 11.9 percent of the scores are below the minimum level, significantly less than in the *No_IQ* condition (p<0.01, Two Sample Test of Proportion, two-sided, n=165); and

do as well as they could. Moreover, they were in competition with other students concerning the feedback process.

also less than in the *Control_IQ* condition ($p<0.01$, Two Sample Test of Proportion, two-sided, n=156). Further evidence of this deliberate "withholding of knowledge" can be inferred from how subjects used the "don't know" option. In the presence of moral consequences, participants chose this option on average 8.90 times, while they chose it only 2.69 times on average in the morally neutral conditions ($p<0.01$, two-sided rank sum test).

We hypothesized that the positive interaction effect in Table 2 originates from the tension between two competing sources of positive self-image, being smart and being moral. One reason to signal positive characteristics to ourselves or others originates from an uncertainty about who we are. If this is true, the interaction effect should be larger for subjects for whom the signal value of performing well or poorly in the IQ test is larger, and vice versa. The intuition is that if test scores are not conveying much information, the relative importance of the signal of being a moral type is higher. This is particularly plausible with respect to the feedback given at the end of the experiment, which informed subjects about whether their test score was above or below the average performance of a large group of other students. If learning about relative performance is a source of positive self-image we would expect that students who are quite certain about their relative standing would gain relatively little from doing well in the test.

We can test this intuition and provide additional evidence in favor of our main hypothesis in using information about study participants' final high-school grade (Abitur).[8] If we run the same regression as in Table 2 but exclude participants with either the highest (top 10 percent, grade strictly better than 1.2) or lowest (lowest 10 percent, grade strictly worse than 2.8) Abitur grades, we find in fact that the coefficient of the interaction term increases by almost 30 percent, from 5.69 in the full sample to 7.39 ($p=0.002$, n=239) in the trimmed sample. While the point estimate for the trimmed sample is considerably higher than the one for the full sample, we note that the two are not statistically significantly different ($p=0.2199$, Wald Test that point estimates are equal).[9] Thus, subjects who derive relatively little value from self-signaling skill may place relatively higher weight on a positive moral self-image.

[8]This information was collected as part of a very short survey taken at the end of the experiment. The Abitur exam is completed at the end of university-track high schools in Germany; passing the exam is a prerequisite for attending university. The final high school grade is the average grade of school achievement for the last two years and therefore informative about cognitive ability. In our sample, the range of grades is between 1.0 (very best) to 3.4, with an average value of 1.96 (SD = 0.57).

[9]Details are available on request.

4 Concluding Remarks

Our findings demonstrate that pursuing pleasures of skill can causally reduce moral behavior. When the same action simultaneously signals one's moral inclination but also competing negative signals concerning skillfulness, moral behavior becomes less likely. Compared to many real life situations, we believe that the decision context implemented in the experiment provides only a lower bound for the relevance of this general trade-off. First, in the Mouse conditions, moral consequences are transparent and unambiguous. Higher levels of performance immediately translate into higher killing probabilities. In addition, subjects were reminded of the consequences, saw the video about gassing mice and were told that mice that were saved would live in appropriate conditions. In contrast, in many real life scenarios, actions and consequences are only indirectly linked, and consequences are often ambiguous or rely on the consent of others. Numerous studies argue that those factors amplify immoral behavior through channels such as moral disengagement (Bandura, 1999), moral wiggle rooms (Dana, Weber, & Kuang, 2007; Serra-Garcia & Szech, 2019), diffusion of responsibility and shared guilt on boards, in committees, teams or between partners (Behnk, Hao, & Reuben, 2017; Falk, Neuber, & Szech, 2020; Huck & Konrad, 2005; Kocher, Schudy, & Spantig, 2018; Rothenhäusler, Schweizer, & Szech, 2018), or market interactions and competition (Falk & Szech, 2013; Kirchler, Huber, Stefan, & Sutter, 2016). Also tournaments in groups can possibly foster the desire to excel even further and potentially increase moral harm. Examples include Charness, Masclet, and Villeval (2014) and Stefan, Huber, Kirchler, Sutter, and Walzl (2020). Second, task performance was not incentivized. Subjects received their show-up fee irrespective of performance and test score. They could easily save mice without any monetary loss. In reality, excelling in challenging tasks is typically associated with instrumental value, such as career benefits, promotions or enhanced public recognition. Third, the signal value of a high test score was only revealed at the individual level. Effects are likely to be stronger if actions evoke social image or reputation effects, in particular if moral consequences are ambiguous. Finally, the desire to signal skillfulness may be more pronounced in tasks, which are of higher self-relevance than succeeding in an IQ-test. Coming back to the initial example on the development of the atomic bomb, part of the identity of being a physicist (or scientist in general) is to acquire new knowledge and develop new techniques. Thus working on and creating new theories or methods in nuclear physics should be highly self-relevant for physicists, more than an IQ-test is for students.

Our findings highlight the existence of a general trade-off between pleasures of skill and moral behavior which apply to many contexts. People like to excel and enjoy pleasures of skill in their domain of expertise, regardless of whether they are managers, bureaucrats, engineers, athletes or scientists. A problem arises if doing a “good job” produces negative externalities, be it managers supporting firms producing morally questionable products, such as chemical weapons or landmines, or engineers constructing machines that systematically deceive users about their ecological consequences. In academia, this trade-off helps explain practices of scientific misconduct, such as faking data. A prominent example is Jan Hendrik Schön, a physicist who has reused the same data set in a number of different experiments, published in top general interest journals; or in the case of social psychologist Diederik Stapel who had received numerous academic awards but was later found to have manipulated and completely made up his own data. In an interview with Brabants Dagblad (October 31, 2011) he admitted that he had failed as a scientist and noted: “I did not withstand the pressure to score, to publish, the pressure to get better in time. I wanted too much, too fast.”

Appendix

A Instructions

In the following we provide an English translation of the instructions for the main two conditions, *IQ* and *NoIQ* (see A.1 and A.2). The two control conditions were identical in wording, respectively, with the only difference that no mouse live was involved. Translations for the two other treatments are available upon request.

A.1 Instructions for the *IQ* treatment

Thank you very much for your participation! For your participation you will receive **15 Euros**. At the end of the experiment you will receive your money in cash.

Please note: Throughout the whole experiment communication between the participants is not allowed. On the computer please only use the functions intended to be used. If you have questions please raise your hand. Your question will then be answered at your cubicle!

Please note: All statements made in these instructions are true. This holds for all experiments carried out by the Bonn Econ Lab, and also for this experiment. In particular, all actions to be taken will be implemented exactly in the way they are described.

15-item Big-5 survey

Test for the Measurement of Intelligence

In a moment you will be taking part in an exercise to measure your intelligence. The test that has been developed for this purpose consists of two parts that are both part of established techniques to measure the intelligence quotient (IQ). Both tests are standardized and frequently used in cognitive psychology.

Generally, intelligence is correlated with many factors of success in a person's life. These comprise, among other things, educational success and average life income.

The first part is a vocabulary test and measures the so-called crystalline intelligence. This test part has no time limit. The second part is a matrix test and measures the so-called fluid intelligence. This consists in turn of two phases, of which the second one has a time limit.

Both components are central devices to measure a person's overall intelligence.

In general: **The more questions you answer correctly, the higher is your measured intelligence quotient.**

The IQ-test you are going to work on in a moment has been conducted already with many students at the University of Bonn, in other experiments and as a part of a questionnaire without any further consequences. At the end of the IQ-test, we are going to tell you how well you performed in comparison to other students. To this end, you are going to learn how many questions on average students have previously answered correctly. You can then compare this value with your result. Your result in the questionnaire has another consequence for a mouse.

Details on the mouse

Figure 2: (Figure shown to subjects in the instructions.)

At issue is a healthy, young mouse. It lives with some other mice together in a small group. The expected lifetime of this mouse is approximately two years. The more questions you answer correctly, the higher is the probability that your mouse gets killed.

For each **correctly** answered question in the test, this probability increases by 0.9 percentage points. The maximal likelihood (if all questions are answered correctly) is 46.8 percent. Hence, if you, for example, answer 30 questions correctly, the probability is 27 percent; for 0 correctly answered questions, it is 0 percent; for 15 correctly answered questions, it is 13.5 percent; for 40 correctly answered questions, it is 36 percent etc.
If the probability is larger than 0 percent, a random generator determines according to this probability whether the mouse dies or lives.

Details on the killing process

If the random generator determines, according to this probability, that the mouse dies, the mouse is gassed. The gas flows slowly into the hermetically sealed cage. The gas leads to breathing arrest. As soon as the mouse is not visibly breathing anymore, it remains in the cage for another 10 minutes. It will then be removed.

Summary

In a moment you can start with the IQ-test. After finishing the tasks, you get to know whether your intelligence result is above or below the average of other students who participated in the test in other experiments. For this purpose, we will tell you your result and the average. The more questions you answer correctly, the higher is the probability that the mouse is killed.

Please press "Continue" when you have read the instructions. *(Button "Continue")*

Video

To visualize the killing of mice by gas, you will in the following see an excerpt of a documentation video (30 seconds). The mouse will be killed in an identical way.

Video

You can now start with the test. The more questions you answer correctly, the higher is the measured intelligence.

A.2 Instructions for the *NoIQ* treatment

Thank you very much for your participation! For your participation you will receive **15 Euros**. At the end of the experiment you will receive your money in cash.

Please note: Throughout the whole experiment communication between the participants is not allowed. On the computer please only use the functions intended to be used. If you have questions please raise your hand. Your question will then be answered at your cubicle!

Please note: All statements made in these instructions are true. This holds for all experiments carried out by the Bonn Econ Lab, and also for this experiment. In particular, all actions to be taken will be implemented exactly in the way they are described.

15–item Big-5 survey

Test

In a moment you will be taking part in another questionnaire. The first part is about guessing words. The second part is about a hands-on assignment of symbols. The answering of the questions in the questionnaire has no further impact for you.

The questionnaire you are going to work on in a moment has been conducted already with many students at the University of Bonn, in other experiments and as a part of a questionnaire without any further consequences. At the end of the questionnaire, we are going to tell you how well you performed in comparison to other students. To this end, you are going to learn how many questions on average the students have previously answered correctly. You can then compare this value with your result. Your result in the questionnaire has further consequences for a mouse.

Details on the mouse (See A.1)

Details on the killing process (See A.1)

Summary In a moment you can start with the questionnaire. After finishing the tasks, you get to know whether your result is above or below the average of other students who participated in the test in other experiments. To this end, we are going to tell you your result and the average. The more questions you answer correctly, the higher is the probability that the mouse is killed.

Please press “Continue” when you have read the instructions. *(Button “Continue”)*

Video (See above.)

Video

You can now start with the test.

References

Ahuvia, A. (2005). Beyond the extended self: Loved objects and consumers' identity narratives. *Journal of Consumer Research*, *32*(1), 171–184.

Ariely, D., Bracha, A., & Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, *99*(1), 544–55.

Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, *3*(3), 193–209.

Behnk, S., Hao, L., & Reuben, E. (2017). Partners in crime: Diffusion of responsibility in antisocial behaviors. *IZA Discussion Paper*.

Belk, R. W. (1988). Possessions and the extended self. *Journal of Consumer Research*, *15*(2), 139–168.

Bénabou, R., Falk, A., & Tirole, J. (2020). Narratives, imperatives, and moral reasoning. *NBER Working Paper Series*, (No. 24798).

Bénabou, R., & Tirole, J. (2005). Self-confidence and personal motivation. In B. Agarwal & A. Vercelli (Eds.), *Psychology, rationality and economic behaviour: Challenging standard assumptions* (pp. 19–57). Palgrave Macmillan UK.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*(5), 1652–1678.

Bentham, J. (1781). An introduction to the principles of morals and legislation. *History of Economic Thought Books*.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972–1059.

Butler, J. V., Serra, D., & Spagnolo, G. (2020). Motivating whistleblowers. *Management Science*, *66*(2), 605–621.

Charness, G., Masclet, D., & Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, *60*(1), 38–55.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory*, *33*(1), 67–80.

Deckers, T., Falk, A., Kosse, F., & Szech, N. (2016). Homo moralis: Personal characteristics, institutions, and moral decision-making. *CESifo Working Paper Series*, (No. 5800).

DellaVigna, S., List, J. A., & Malmendier, U. (2012). Testing for altruism and social pressure in charitable giving. *Quarterly Journal of Economics*, *127*(1), 1–56.

Dittmar, H., & Pepper, L. (1992). Materialistic values, relative wealth and person perception: Social psychological belief systems of adolescents from different socio-economic backgrounds [Association for Consumer Reserach and School of Business, Provo UT]. *Meaning, Measure, and Morality of Materialism*, 40–45.

Ewers, M., & Zimmermann, F. (2015). Image and misreporting. *Journal of the European Economic Association*, *13*(2), 363–380.

Falk, A. (2018). Facing yourself - a note on self-image. *Working Paper*.

Falk, A., Neuber, T., & Szech, N. (2020). Diffusion of Being Pivotal and Immoral Outcomes. *The Review of Economic Studies*.

Falk, A., & Szech, N. (2013). Morals and markets. *Science*, *340*(6133), 707–711.

Feynman, R. P. (1985). *''surely you're joking mr. feynman!" adventures of a curious character*. WW. Norton Company, Inc.

Fischbacher, U. (2007). Z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Furby, L. (1991). Understanding the psychology of possession and ownership: A personal memoir and an appraisal of our progress. *Journal of Social Behavior and Personality*, *6*(6), 457–463.

Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: The results of a large-scale field experiment. *American Political Science Review*, *102*(1), 33–48.

Gert, B., & Gert, J. (2017). The definition of morality. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Fall 2017). Metaphysics Research Lab, Stanford University.

Greiner, B. (2004). An online recruitment system for economic experiments. In K. Kremer & V. Macho (Eds.), *Forschung und wissenschaftliches Rechnen. GWDG Bericht 63.* (pp. 79–93). Ges. für Wiss. Datenverarbeitung.

Huck, S., & Konrad, K. A. (2005). Moral Cost, Commitment, and Committee Size. *Journal of Institutional and Theoretical Economics (JITE)*, *161*(4), 575–588.

Kirchler, M., Huber, J., Stefan, M., & Sutter, M. (2016). Market design and moral behavior. *Management Science*, *62*(9), 2615–2625.

Kocher, M. G., Schudy, S., & Spantig, L. (2018). I lie? we lie! why? experimental evidence on a dishonesty shift in groups. *Management Science*, *64*(9), 3995–4008.

Kuhnen, C. M., & Tymula, A. (2012). Feedback, self-esteem, and performance in organizations. *Management Science*, *58*(1), 94–113.

Lacetera, N., Macis, M., & Slonim, R. (2012). Will there be blood? incentives and displacement effects in pro-social behavior. *American Economic Journal: Economic Policy*, *4*(1), 186–223.

Lifton, R. J., & Mitchell, G. (1995). *Hiroshima in america: Fifty years of denial.* G.P. Putnam's Sons.

Loewenstein, G. (1999). Because It Is There: The Challenge of Mountaineering - for Utility Theory. *Kyklos*, *52*(3), 315–343.

Marshall, A. (2009). *Principles of economics: Unabridged eighth edition.* Cosimo, Inc.

Mårtensson–Pendrill, A.-M. (2006). The manhattan project - a part of physics history. *Physics Education*, *41*(6), 495–496.

Mazar, N., Amir, O., & Ariely, D. (2008). More ways to cheat-expanding the scope of dishonesty. *Journal of Marketing Research*, *45*(6), 651–653.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77.

Norton, M. I., Mochon, D., & Ariely, D. (2012). The ikea effect: When labor leads to love. *Journal of Consumer Psychology 22*, (3), 453–460.

Prelec, D., & Herrnstein, R. J. (1991). Preferences or principles: Alternative guidelines for choice. *Strategy and choice* (p. 319). MIT Press.

Rothenhäusler, D., Schweizer, N., & Szech, N. (2018). Guilt in voting and public good games. *European Economic Review*, *101*, 664–681.

Russell, B. (1960). The social responsibilities of scientists. *Science*, *131*(3398), 391–392.

Serra-Garcia, M., & Szech, N. (2019). The (in)elasticity of moral ignorance. *CESifo Working Paper Series*.

Stefan, M., Huber, J., Kirchler, M., Sutter, M., & Walzl, M. (2020). Monetary and social incentives in multi-tasking: The ranking substitution effect. *MPI Collective Goods Discussion Paper*, (2020/10).

Sternberg, R. J., Grigorenko, E. L., & Bundy, D. A. (2001). The predictive value of iq. *Merrill-Palmer Quarterly*, *47*(1), 1–41.

Weaver, W., Snow, C. P., Hesburg, T. M., & Baker, W. O. (1961). The moral un-neutrality of science. *Science*, *133*(3448), 255–262.

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, *66*(5), 297–333.