CENTER FOR
ECONOMIC
BEHAVIOR &
INEQUALITY

CEBI WORKING PAPER SERIES

Working Paper 12/24

SORTING AND WAGE PREMIUMS IN IMMORAL WORK

Florian H. Schneider

Fanny Brun

Roberto A. Weber

ISSN 2596-447X

CEBI
Department of Economics
University of Copenhagen
www.cebi.ku.dk

Sorting and wage premiums in immoral work

Florian H. Schneider, Fanny Brun and Roberto A. Weber[*]

May 13, 2024

We use surveys, laboratory experiments and administrative data to study how heterogeneity in the perceived immorality of work and in workers' aversion to acting immorally impact labor market outcomes. Immoral work is associated with higher wages, both in administrative data and in causal experimental evidence. Individuals more willing to engage in immoral conduct find employment in firms and industries perceived as immoral less aversive and have higher employment rates in immoral work in the laboratory. These phenomena appear to be driven by impure social motives, reflecting a desire not to be involved with immoral work, rather than by consequentialist concerns.

Keywords: Wage premium, immoral behavior, impure social preferences, sorting, experiments

JEL Codes: C92, J31, D03

[*] Schneider: Department of Economics and Center for Economic Behavior and Inequality (CEBI), University of Copenhagen, Copenhagen, Denmark, florian.schneider2@econ.uzh.ch. Brun: independent researcher, fannybrun@protonmail.com. Weber: Department of Economics, University of Zurich, Zurich, Switzerland, roberto.weber@econ.uzh.ch. We thank participants at several conferences and seminars for helpful comments and suggestions. We also gratefully acknowledge the financial support of the Swiss National Science Foundation (Grant No. 100018_165943) and the Danish National Research Foundation (DNRF134).

1. Introduction

There exists a widespread impression that working in some professions, organizations, and industries—e.g., tobacco or weapons—is inherently immoral, and that workers' aversion to performing such "immoral work" may impact labor market outcomes. This perspective dates back to Adam Smith (1776; Book I; Ch. X), who wrote about professions then perceived as irreputable, "The exorbitant rewards of players, opera-singers, opera-dancers, etc., are founded upon [..] the discredit of employing them in this manner. [..] Should the public opinion or prejudice ever alter with regard to such occupations, their pecuniary recompense would quickly diminish. More people would apply to them, and the competition would quickly reduce the price of their labour." That is, the perceived immorality of a line of work may reduce the supply of available workers, attracting those who care the least about acting immorally and yielding a wage premium, similarly to other ways in which the aversiveness of work can yield compensating differentials (Rosen, 1986).[1]

However, despite the intuitive appeal of such relationships, little empirical evidence links heterogeneity in the perception that work is immoral with individuals' willingness to perform such work and to the resulting labor market outcomes. In this paper, we provide novel evidence testing such relationships, using a combination of surveys, laboratory experiments and administrative labor market data. The administrative data provide the clearest evidence of the economic relevance of these relationships, but in these data industries perceived as immoral might differ in unobservable aspects from other industries, making it difficult to establish a causal relationship. The control provided by laboratory experiments allows us to observe what outcomes arise as the nature of work changes *only* in the extent to which it involves doing things generally perceived as immoral and to investigate underlying mechanisms. We additionally use evidence from surveys to obtain insights into relationships between individuals' concerns for morality and their willingness to work in varied real-world firms and industries. Thus, our paper highlights the value of employing complementary research methods to address complex economic phenomena.

Our work focuses on two hypotheses from simple theoretical analysis of how individuals' heterogeneous aversion to performing immoral acts may interact with jobs that vary in their perceived immorality. The hypotheses reflect the above relationships: first, that work perceived as

[1] There is anecdotal evidence that firms operating in industries perceived as immoral face recruitment challenges. For example, following the Cambridge Analytica scandal, Facebook struggled to attract top talent (CNBC, 2019). Tobacco companies deem difficulties in recruitment arising from their image as sufficiently important to warrant disclosure to regulators and shareholders (British American Tobacco, 2015, p. 37; Philip Morris International Inc., 2015, p. 14).

immoral yields higher wages—as long as workers care enough about avoiding immoral work—and, second, that immoral work attracts those workers least concerned with immoral conduct.[2]

Earlier research, reviewed in the next section, documents that work with positive social impact may attract more pro-socially oriented workers willing to work at a discount. This seems consistent with the relationships that we investigate—although reflecting a preference for moral work, rather than distaste for immoral work—and perhaps yielding a mirror image of our hypothesized results. However, it is not clear that such evidence yields straightforward predictions for what outcomes to expect in markets for immoral work. For instance, the aversion to immoral work and the attraction to moral work may differ in their distributions and intensities, resulting in different impacts on outcomes. More importantly, jobs in industries perceived as immoral need not be aversive to those pursuing positive social impact. A worker concerned with limiting the harm produced by tobacco or firearms may have sizable opportunities to do so by working in these industries, e.g., by designing safer products or responsible marketing. Thus, if workers are mainly motivated by a consequentialist desire for positive impact, markets for moral and immoral work may both yield sorting in by pro-social types and wage discounts.[3] However, as we discuss in detail and show theoretically later, non-consequentialist (or, "impure") moral preferences—as with social image concerns or warm glow—may produce very different predictions for moral and immoral work, even when the potential for social impact is identical. That is, "moral" types may find working in "immoral" industries aversive, even if they could do "good" by working there.

Our results provide support for our primary hypotheses, both in and out of the laboratory. Table 1 provides an overview of our main findings.

First, we show that work perceived as immoral commands a wage premium over comparable work not perceived as immoral. Combining novel ratings of the immorality of several industries with administrative labor market data (Section 4), we find that industries generally perceived as immoral yield higher wages, even after controlling for many observable worker, job, and industry characteristics. Moreover, in two laboratory labor markets—which vary by treatment

[2] We use "immoral work" to refer to employment in organizations or industries generally perceived as immoral and distinguish this from immoral work activities within firms or industries not perceived as immoral, which is not our focus. Furthermore, individuals hired for "immoral work" may not necessarily perform activities that are inherently immoral or harmful. We return to this distinction throughout the paper, including as a focus of our second experiment.

[3] Recent evidence suggests that firms operating in "green" (environmentally friendly) industries have substantially less scope to further improve environmental impacts than firms in "brown" industries (Hartzmark and Shue, 2023), suggesting that workers or investors aiming to create positive impacts might target their efforts toward the latter.

only whether employment requires doing something generally perceived as immoral—we observe a causal relationship indicating that wages are persistently higher for such immoral work (Sections 5 and 6). These wage premiums are substantial and are not eliminated by market experience.

Table 1. Overview of our evidence on wage premiums and sorting

	Laboratory labor markets	*Labor markets outside the laboratory*
Immorality premium	Causal evidence for a wage premium for immoral work (Section 5, Figure 4; Section 6, Figure 5)	Correlation between perceived industry immorality and wages (Section 4; Figure 1)
Sorting	Immoral types are more likely to be hired, but only for immoral work (Section 5; Table 2; Section 6, Table 3)	Immoral types state a greater willingness to work in firms and industries others perceive as immoral (Section 7; Figure 6)

Second, we provide evidence of sorting by "immoral" types into "immoral" work, both in the laboratory and the field. We measure individuals' aversion to acting immorally outside a market context. This measure predicts individual labor market outcomes. In laboratory markets (Sections 5 and 6), immoral types are employed more frequently, but only when work involves immoral acts. In our survey data (Section 7), immoral types report a significantly greater willingness to work in firms and industries that are generally perceived by others as immoral.

Our two laboratory experiments provide corroborative evidence for our main hypotheses, but they differ in important ways. Both studies involve labor markets in which workers state reservation wages for being hired to perform a task, and wages are determined by the resulting labor supply and exogenously determined labor demand. Both studies contain "immoral" work conditions, in which employment involves making a dishonest statement and taking an action that reduces donations to organizations engaged in positive social impact,[4] and introduce social image concerns by disclosing workers' employment. Both experiments also include a neutral work condition, in which workers perform virtually identical activities but with no dishonesty or harm.

However, the impact of individuals' market behavior differs between the two studies. In Study 1, it is impossible for workers employed in immoral work to produce consequential positive

[4] Deception and harm are often associated with "immoral" firms and industries. Tobacco companies have long been accused of engaging in misleading marketing regarding smoking's harmful effects (Heath, 2016). Similar attention has recently focused on the role of the aggressive marketing of assault weapons in facilitating widespread gun violence (Karni, 2022). The financial industry, regularly confronted with perceptions of immorality, is also often associated with deceptive practices that hide risk and harm public finances (Akerlof and Shiller, 2015).

impact; positive impact is only possible by withholding one's labor supply. Thus, the sorting and wage premiums observed may reflect moral individuals' desire for positive impact and willingness to pay a cost for doing so, as when workers accept lower wages for socially responsible work.

Study 2 additionally contains a "moral" work condition, which involves making an honest statement and generating positive social impact analogous to the negative one in the immoral work condition. Critically, our second study holds fixed the quantity of labor employed, meaning that individual workers cannot influence total employment for either moral or immoral work. Study 2 also provides workers with discretion to reduce the harm produced by their employment (or, for moral work, to increase the benefit) at a cost. Thus, a worker in Study 2 motivated to produce positive social impact cannot influence total employment by reducing their labor supply for immoral work, but can generate positive impact by obtaining employment. Purely consequentialist "moral" workers should thus seek employment in the immoral work condition of Study 2 at the same rate as for moral work. Instead, we find that morally motivated workers reduce their labor supply in the market for immoral work, consistent with non-consequentialist motives. We thus provide evidence that the aversion to employment in immoral work does not simply reflect a preference driving individuals to seek employment in jobs to have positive social impact, but instead that there are unique costs associated with working in jobs perceived as immoral.

Our findings have potentially important implications for understanding the nature and consequences of moral preferences in markets, which we discuss further in the next section and in the Conclusion. For example, our second study sheds light on the degree to which individuals' moral concerns are motivated by consequentialist considerations, like a desire for social impact, or non-consequentialist motives, like a desire for positive social image. The nature of such concerns may be critically important for the equilibrium impacts of moral preferences in markets (Dewatripont and Tirole, 2023; Köszegi and Kaufman, 2023).

Moreover, the evidence we provide of sorting suggests that if industries with greater potential for social harm—e.g., weapons manufacturing—are perceived as more immoral, they may consequently attract and reward those individuals least concerned with minimizing negative social impacts. Our second experiment provides evidence on the consequences of such sorting, by allowing us to observe the degree of social impact produced in markets for moral and immoral work. Surprisingly, we find that workers more often reduce harm when hired for immoral work than increase benefit when hired for moral work. This unexpected finding seems particularly puzzling given that we observe sorting into immoral work by those types whose earlier behavior indicates less concern for acting morally. While we lack definitive evidence for why this occurs, we cautiously propose three possible interpretations based on income, moral licensing, and moral cleansing effects, and provide suggestive evidence for income effects. Additionally, we demonstrate that a simple extension of the model that we use to motivate our investigation can encompass these effects, successfully capturing all the results in this paper.

The rest of our paper proceeds as follows. In the next section, we discuss how this paper relates and contributes to previous literature. Section 3 describes a simple theoretical framework that we use to develop hypotheses. In Section 4, we use Swiss labor market data to investigate the relationship between the perceived immorality of work and wages. Sections 5 and 6 present the design and results of our two laboratory experiments, while Section 7 describes our survey study. Section 8 concludes by discussing implications of our findings for policy.

2. Related Literature

There is considerable evidence that people exhibit heterogeneous concerns for moral conduct, including in market contexts (Abeler, Nosenzo and Raymond, 2019; Fehr and Charness, 2023). However, there are also reasons to believe that the impacts of moral concerns may be mitigated in competitive markets (Dufwenberg, et al., 2011; Bartling, Weber and Yao, 2015; Kirchler, Huber, Stefan and Sutter, 2016; Ziegler, Romagnoli and Offerman, 2020). Such concerns might particularly apply to labor markets, where the preferences of the marginal worker set wages and where repeatedly forgoing profitable job opportunities may erode workers' moral motives. On the other hand, moral considerations might be particularly relevant in labor markets, as job choices are typically visible to others and may significantly impact workers' sense of identity and social image (Bénabou and Tirole, 2011; Oh, 2021). Our studies add further evidence to better understand how moral concerns affect market behavior and outcomes. Our observation of substantial wage premiums for immoral work—both inside and outside of the laboratory—suggests that many workers are willing to forgo financial gains to avoid work in immoral industries. Moreover, our study sheds light on the degree to which market outcomes reflect consequentialist or non-consequentialist moral motives, a distinction important for understanding the impacts of moral preferences in markets (Dewatripont and Tirole, 2023; Köszegi and Kaufman, 2023).

There is related evidence on sorting by pro-social "mission-oriented" types into the public sector (Carpenter and Myers, 2010; Dur and Zoutenbier, 2014; Fisman et al, 2015; Hanna and Wang, 2017; Ashraf, et al., 2020; Barfort, et al., 2019; Friebel, Kosfeld and Thielmann, 2019).[5] Moreover, several studies investigate whether nonprofit employees earn less than for-profit employees (e.g., Leete, 2001; Mocan and Tekin, 2003).[6] Recent work by Hedblom, Hickman and List (2019) finds that a firm advertising its work as socially oriented, substantially increases the number of applicants. Relatedly, Hu and Hirsh (2017) show that people are willing to accept lower salaries for more meaningful work. Krueger, Metzger and Wu (2021) find that sectors and firms with better environmental ratings pay lower wages, even when keeping the occupation fixed and controlling for a rich set of worker characteristics. We complement this work by exploring variation in the perceived *immorality* of work and the aversion to acting immorally—which is potentially distinct from the desire to have positive impact by accepting "moral work"—as the driving sources of heterogeneity. Moreover, our second laboratory experiment considers both immoral and moral work, thereby investigating the importance of non-consequentialist moral motives for understanding differences in labor market outcomes for moral and immoral work.

Related evidence on compensating differentials documents wage premiums for presumably immoral behavior, like prostitution (Arunachalam and Shah, 2008), and for work in professions and firms with low levels of perceived social responsibility (Frank, 1996). This work shares features with our correlational evidence in Section 4, but uses narrower samples. Correlational evidence like this might result from other unobserved worker and job characteristics. For instance, it is unclear whether compensation for work like prostitution is for perceived immorality or other aversive job features (Edlund and Korn, 2002; Gertler, Shah and Bertozzi, 2005).[7] Such studies

[5] There is also correlational evidence that people in specific industries differ in their values (Ashraf, Bandiera and Delfino, 2020) and trustworthiness (Gill, Heinz, Schumacher and Sutter, 2020). Carter and Irons (1991) find that economics students exhibit lower concerns for morality; however, this study does not document that these moral concerns drive differential selection into different kinds of work rather than the opposite relationship (Frank, Gilovich and Regan, 1993; Cohn, Fehr and Maréchal, 2014; Ashraf and Bandiera, 2017). Our study also relates to research on effort and sorting by mission-oriented types (Besley and Ghatak, 2005; Delfgaauw and Dur, 2008; Ariely, Bracha and Meier, 2009; Fehrler and Kosfeld, 2014; Tonin and Vlassopoulos, 2015; Cassar and Meier, 2018; Dur and van Lent, 2019), though this research focuses on worker motivation and effort within firms. Our work also relates to theoretical studies on the allocation of skilled labor to sectors more or less important for society (Murphy, Shleifer and Vishny, 1991), and how tax policy can reallocate labor to socially valuable industries (Lockwood, Nathanson and Weyl, 2017).

[6] Early studies yielded mixed correlational evidence, likely due to methodological challenges in estimating compensating wage differentials using observational data (see the discussion in Mas and Pallais, 2017). Recent papers on compensating differentials (for non-moral factors) instead rely on experimental methods and/or stated preferences (Carpenter, Matthews and Robbett, 2017; Mas and Pallais, 2017; Wiswall and Zafar, 2018), as we also do here.

[7] Related work in finance (Hong and Kacperczyk, 2009) demonstrates that investing in firms that engage in immoral activities ("sin stocks") yields higher returns. However, other industry and firm characteristics, such as litigation risk, may also differ for these types of investments (Blitz and Fabozzi, 2017).

also fail to measure workers' heterogeneous concerns for morality—an important feature of our study—as a key driver of the relationship.

3. Theoretical framework

Our study is guided by hypotheses from a simple labor market model with variation in perceived work immorality and in workers' heterogeneous concern for avoiding immoral acts.[8] We present the detailed model and analysis in Appendix C.1; here, we focus on the model's key predictions.

We examine a labor market for a job, $j \in J$, which might involve doing work perceived as immoral. The perceived immorality of j is measured by a function $I: J \to [0, \infty)$, where $I(j') > I(j)$ means that job j' is perceived as *more immoral* than job j. Firms decide whether to hire a worker to do j at the market wage, *w*, and workers decide whether to accept work j for the market wage. Workers differ in their concern with avoiding immoral work, $\theta_i \geq 0$. A worker of type θ_i accepts job j if the utility from doing so is higher than that of an outside option, or

$$u_i^{accept}(j, w) = w - c - \theta_i * I(j) \geq \underline{u},$$

where $c \geq 0$ is the worker's cost of effort and $\underline{u} \geq 0$ the workers' reservation utility. We focus on jobs that differ only in perceived immorality and, therefore, assume that c is independent of j.[9] While our laboratory experiments are designed to satisfy this assumption, it may be violated in labor market data, as we discuss later. We assume that workers are not concerned with the consequential harm produced by immoral work, but rather have an aversion to personally implementing work perceived as immoral. For example, θ_i can be understood as a reduced-form representation of social image concerns in a signaling model (Bénabou and Tirole, 2006; DellaVigna, List, Malmendier and Rao, 2016) or as impure altruism (Andreoni, 1989).

Our two results derive the primary hypotheses for our analysis. The first result (Proposition 1) shows that there is an immorality premium for immoral jobs: an increase in the immorality of a job, $I(j)$, decreases labor supply and thus increases the equilibrium wage. However, this wage

[8] Our framework is a simplification of earlier models on compensating wage differentials (e.g., Rosen, 1986). We do not seek to expand this literature, but rather apply it to a context where the relevant job dimension is immorality. Unlike most models of compensating wage differentials, we do not have multiple labor markets, but instead one and a fixed outside option, which is consistent with the design in our first laboratory experiment. This abstraction simplifies both the theory and the experiment. Appendix C.2 shows that our results also hold in a model with two types of jobs, an immoral job and a neutral job, and our second laboratory experiment considers such a setting with two jobs.

[9] Note that while we simplistically refer to c as the cost of effort, it can represent any job attributes impacting workers' utility that are independent of the immoral nature of the job, $I(j)$.

premium is insignificant if workers care little about morality (Corollary). Our second main result (Proposition 2) states that those individuals least concerned with avoiding immoral work—i.e., those with low θ_i—sort into accepting immoral jobs, while those more concerned with morality refuse to do the job for the equilibrium wage. That is, wage premiums arise precisely because those who find immoral work most distasteful opt out of such jobs.

4. Evidence of an immorality premium in the Swiss labor market

We test for wage premiums in industries perceived as involving immoral work. To do so, we obtain novel measures of the perceived immorality of various industries in Switzerland, which we compare with wages from the Swiss Labor Force Survey (SLSF), a representative worker sample compiled by the Swiss Federal Statistical Office. We study whether perceived industry immorality can account for a portion of wages unexplained by observable worker and industry characteristics.

Our general approach for obtaining ratings of perceived industry immorality proceeds in two steps: (i) selection of a broad set of industries and (ii) independent ratings of perceived industry (im)morality. We employ different methods for these two steps.

In our first method, we identified industries that we (all three authors) jointly perceived as involving work activities likely to be widely seen as immoral; we did so before looking at any data from these industries, including wages (for details, see Appendix B).[10] We also chose five comparison industries from within the same industrial branch with similar distributions of education levels and nine additional industries representing large shares of employment in Switzerland. We then obtained independent ratings of the perceived (im)morality of these industries through a survey of 177 university students in Switzerland.[11] We interpret this variable as a measure of the perceived immorality of working in industry j, or $I(j)$, a key component of our theoretical analysis. The horizontal axis of Figure 1a shows the mean ratings for each industry.

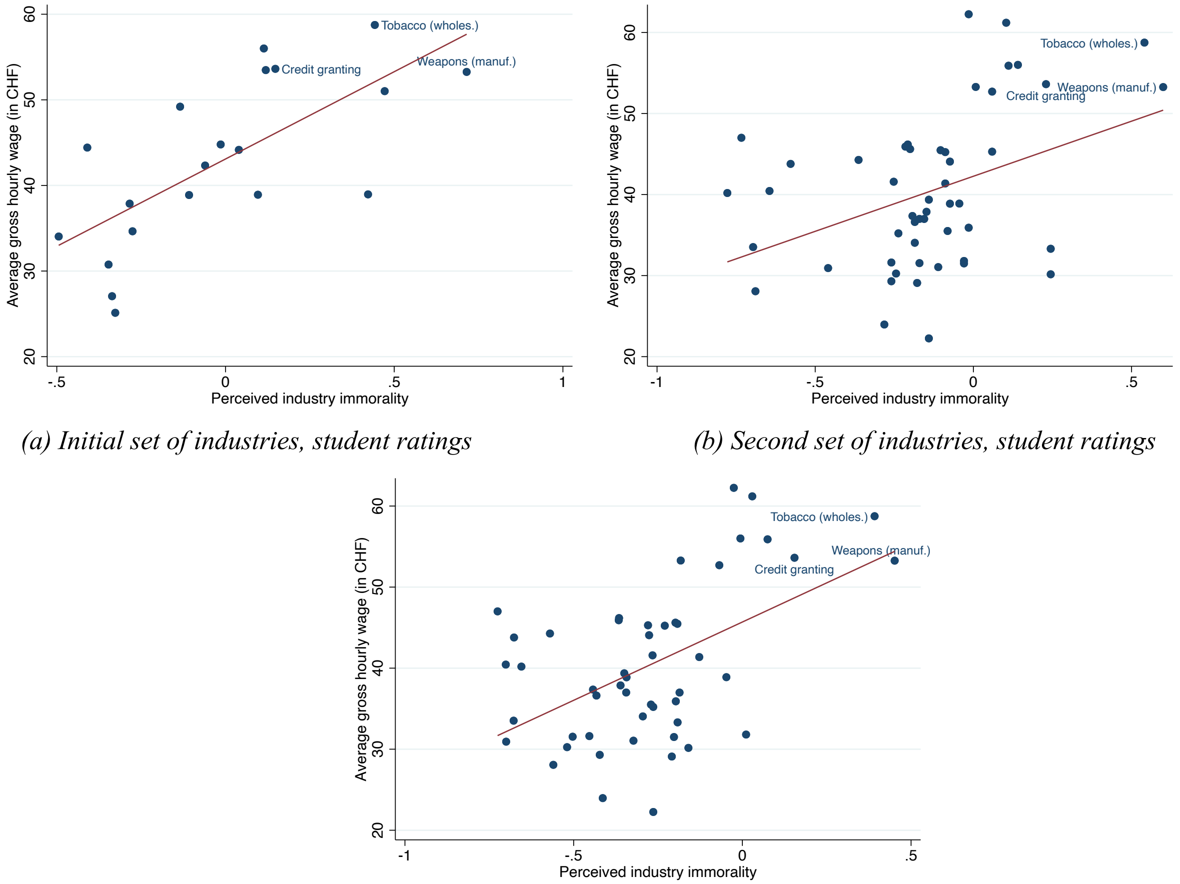
While we did not look at wages when selecting the industries, a natural concern is that choosing the sample of industries ourselves possibly (unconsciously) biases the sample toward those likely to confirm our hypothesis. Hence, we also implemented step (i) by asking research

[10] This yielded six "immoral" industries: gambling and betting, monetary intermediations, credit granting, manufacture of tobacco, wholesale of tobacco, and manufacture of weapons and ammunition. These include industries regularly classified as "sin industries" in financial research (Hong and Kacperczyk, 2009; Blitz and Fabozzi, 2017).

[11] Students on the campuses of the University of Zurich (UZH) and the Federal Institute of Technology (ETH) rated each industry on a 5-point Likert scale ("very moral" (-1) to "very immoral" (1)). We averaged the responses. These survey data were collected as part of our survey studies, which we describe in more detail in Section 7.

assistants unfamiliar with the research question to identify 50 industries (for details, see Appendix B).[12] For step (ii), we collected immorality ratings for this second set of industries with two samples: a new sample of students (N=45) and a larger survey sample broadly representative of the German- and French-speaking populations of Switzerland (N=303). The resulting industry ratings are shown on the horizontal axes of Figures 1*b* and 1*c*. The overall immorality ratings are similar (correlation = 0.91), supporting the stability of immorality perceptions across populations.

Figure 1: Correlation between wages and perceived industry immorality



(a) Initial set of industries, student ratings

(b) Second set of industries, student ratings

(c) Second set of industries, ratings from representative sample

Source: Weighted data from the SLFS, years 2010-2016 (wage) and our own survey (perceived industry immorality). Notes: Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Real gross hourly wage in 2010 CHF (1 CHF ≈ 1 US Dollar). N = 32,638 for panel (a) and N = 47,935 for panels (b) and (c). We label the three industries that both the research assistants and we expected to be perceived as immoral (Manufacture of weapons and ammunition, Wholesale of tobacco products, Credit granting).

[12] This yielded five "immoral" industries: manufacture of weapons and ammunition, wholesale of tobacco, processing and preserving of meat (except poultry meat), credit granting and processing and preserving of poultry meat. We thank Uri Gneezy for suggesting this approach.

The vertical axis of Figure 1 plots the mean real gross hourly wage (in 2010 Swiss Francs) in each industry. These data are the reported hourly wages of employees surveyed in the SLFS. We use data from the 2010 to 2016 waves. The sample includes 32,638 observations for panel (a), and 47,935 observations for panels (b) and (c). The strong positive relationship in each panel supports the hypothesis that industries with greater perceived immorality are associated with wage premiums, which holds across different ways of selecting industries and different populations from which we elicit the ratings of perceived immorality.

Of course, the relationships in Figure 1 ignore individual worker characteristics, which vary across industries, and other industry characteristics that may partially account for wage gaps. To partially address this concern, the SLSF data allow us to control for various worker and job characteristics (age, gender, nationality, experience, full-time equivalent, managerial duties, employer location fixed effects). In addition, we match the SLSF data with industry characteristics from other sources (industry size from STATENT and industry sales from Value Added Tax Statistics). We then regress the natural logarithm of real gross hourly wages on perceived industry immorality, controlling for worker, job and industry characteristics (using OLS).

The results (shown in Appendix Table A1) show large and statistically significant wage premiums. Using the initial set of industries, we estimate that a one standard deviation increase in perceived industry immorality is associated with a 14.8 percent higher (geometric) mean hourly wage ($z = 3.59$, $p < 0.001$, 95%-CI: [6.7, 22.9]).[13] If we instead use the second set of industries, the wage premium is 7.4 percent ($z = 2.67$, $p = 0.008$, 95%-CI: [2.0, 12.9]) when using student ratings and 9.1 percent ($z = 4.66$, $p < 0.001$, 95%-CI: [5.3, 13.0]) when using ratings from the representative sample. While the estimated wage premiums vary somewhat when using different industries and obtaining ratings from different samples, they are always large and highly statistically significant. In Appendix Tables B2 and B4, we show that the finding of an immorality wage premium is robust to the use of varying sets of controls and the absence of any controls. Moreover, Appendix Table B6 shows that we find similar results if, instead of measuring $I(j)$, we focus on a set of "sin industries", thereby following the approach used in the financial literature to study returns to "sin stocks" (e.g., Hong and Kacperczyk, 2009; Blitz and Fabozzi, 2017).

[13] We obtain this number by doing the following calculation: $e^{0.138} - 1 \approx 0.148$. We use the delta method to calculate the corresponding z-values, p-values and confidence intervals (CIs).

Our analysis thus demonstrates wage premiums for work perceived as immoral, consistent with our first hypothesis. However, the correlational aspect of the relationship leaves open the possibility that additional unobserved industry characteristics may explain the observed relationships. That is, our theoretical framework rests on the assumption that work characteristics unrelated to immorality (in our model, c) are independent of industries' perceived immorality, $I(j)$. Moreover, the analysis does not tell us whether the workers employed in these industries differ in their moral concerns. In the following three sections, we explore the model's predictions more carefully—controlling for unobservable aspects of work, thereby keeping c constant, and making a clearer connection to subjects' heterogeneous concerns for morality.

5. Study 1: Sorting and wage premiums in a laboratory labor market

Our first study investigates the preferences and behavior of subjects in the role of workers in a labor market. It consists of an online questionnaire, followed by a laboratory session approximately one week later. The questionnaire measures subjects' preferences regarding future employment possibilities. Appendix Figure A1 provides an overview of the sequence of the study. We discuss the online questionnaire in detail in Section 7, where we also report the analysis of the resulting data. In this section, we focus on the design and results of the laboratory experiment.

In the laboratory, we first elicit a measure of concern for morality (θ in the theoretical framework) using an incentivized behavioral task. Subjects then participate in a labor market for 15 periods, in each of which they submit reservation wages and are potentially hired based on their and other workers' wage requests and an automated demand schedule.

Our experiment exogenously varies *only* whether work involves doing something generally perceived as immoral, to investigate the causal impact on labor market outcomes. We design an "immoral" act akin to giving fraudulent financial advice to a non-profit organization, thereby harming the non-profit's ability to provide aid. Deceptive communication and social harm are features of many jobs perceived as immoral. We inform subjects that each session is endowed with an initial donation to a UNICEF fund providing malaria treatments for children (the aid recipients).[14] However, the actual final donation is affected by participants' behavior. Specifically,

[14] To strengthen the moral context and link the UNICEF money to "donors," we link these donations to donations generated by third parties. Prior to the laboratory sessions, we approached individuals who donated blood as part of a donation campaign and asked if they would agree that we potentially make a donation to UNICEF as a complement to their blood donation. Most donors agreed. We did not otherwise obtain information or choices from these donors.

subjects in our experiment are hired to provide written advice to a "client" (a subject participating in a subsequent survey and serving a role akin to a non-profit employee making financial decisions on its behalf). We vary, by treatment, whether workers are assigned to a market with *neutral* jobs that involve honest advice with little impact on the client and the UNICEF donation or to a market with *immoral* jobs that involve dishonest advice that harms the client and the donation.

After the laboratory sessions, we recruit a separate sample of individuals at public locations to serve two functions. First, they fill the role of "clients" who receive written recommendations from laboratory subjects and act on this advice. From these choices, the clients earn money and determine the size of the UNICEF donation. Second, these participants complete a survey in which they rate the degree to which various industries and firms are "moral" or "immoral." We use these ratings as one approach to construct perceived industry immorality in the analysis in Section 4.

5.1 Design details

Laboratory sessions consist of 24 participants. Before entering the lab, participants pose for a neutral portrait photograph that we use to make labor market outcomes public.[15] Participants read an information sheet at the beginning of the session about the consequences of malaria for children and the need for treatments—we adopt wording from public UNICEF materials. Participants receive instructions both on paper and with pre-recorded audio files.

5.1.1 Behavioral measure of concern for morality (θ)

Participants first play an incentivized game that measures their willingness to lie for personal gain while causing others harm, in a non-market environment (Gneezy, Rockenbach and Serra-Garcia, 2013). In the game, Participant A privately observes a computerized die roll (r) and sends a message with a reported number (m) to Participant B. Participant A may claim m to be any integer from 1 to 6, regardless of r, receiving a payoff of $5+m$ CHF (1 CHF ≈ 1 USD).[16] This provides an incentive to lie whenever r is less than 6. Participant B then decides whether "to follow" or "not

[15] This mirrors labor markets outside the laboratory, where one's workplace is often observable, and reflects the impure motives such as social image concerns that play a key role in our theoretical analysis. Our design likely induces weak social image concerns. In a separate study conducted by one of the researchers using the same subject pool, participants saw pictures of other participants in the session and were asked if they knew them. In 99% of cases, participants did not recognize the person, suggesting that the likelihood of not knowing another participant in one's market is around 0.99^5 = 95%. Considering the large number of students in the subject pool, it is unlikely that participants would encounter and recognize each other after the experiment.

[16] The instructions and interface referred to earnings in "points," which we converted to money at the rate of 20 points = 1 CHF. For clarity, we present the design and results in terms of ultimate payments in Swiss francs (CHF).

to follow” Participant A’s message. Not following yields Participant B 1.5 CHF and the donation to UNICEF remains unaffected. If Participant B follows the message and $m = r$, Participant B earns 5 CHF and the initial donation to UNICEF increases by an amount corresponding to one anti-malarial treatment.[17] However, if Participant B follows the message and $m \neq r$, Participant B earns no money and the donation *decreases* by one treatment. We inform participants that, at the end of the session, their decisions as Participant A will be publicly displayed to other participants in the session, along with their portrait photograph.

Every participant initially plays as Participant A. We use the strategy method to elicit Participant A’s message for every possible die roll. At the end of the laboratory session, 5 of the 24 participants in the session have their role changed from Participant A to Participant B. These Participants B are then matched with five of the remaining Participants A and decide whether to follow the corresponding message. All participants whose role is not switched are paid based on their choice as Participant A, independently of whether they are matched with a Participant B.[18]

5.1.2 Experimental labor market

In the labor market, subjects play the role of workers competing to be hired by automated firms. Subjects receive general instructions about the labor market and then answer comprehension questions. Subjects then receive information about the nature of the job, which varies by treatment, and answer a new set of comprehension questions about the job.

The job. In both conditions, workers can be hired as an “advisor” whose job is to advise another participant outside the laboratory, the “client.” The advisor must provide a written recommendation to select one among ten choice options (labeled “A” through “J”; see Appendix Table A2).[19] Nine options increase the client’s reward by 1 CHF and increase the donation to UNICEF by an amount corresponding to the treatment of one child. However, there is always one option that gives 0 CHF to the client and *reduces* the UNICEF donation by one treatment.

[17] The actual cost of providing 30 malaria treatments, according to UNICEF, was CHF 29. To create small units with a tangible moral component, our instructions always referred to the amount corresponding to treating “one child.”

[18] This implies that Participant As (whose role was not switched) receive their own payment with certainty. Their decision, however, only has consequences for Participants B (and UNICEF) with a probability of 26.3 percent. This corresponds, roughly, to the stochastic impacts in our experimental labor market.

[19] Study 1 involves a single labor market with only one type of job (either immoral or neutral, depending on treatment condition); a worker not hired does no work. We chose this abstraction to simplify the experiment. Labor market behaviors may differ in situations where participants have the option to choose between two job types, such as an immoral and a morally neutral job. We address this potential limitation in Study 2.

The client receives the advisor's recommendation, and then selects one of the ten options. The client only knows that the selected option determines the client's payment for completing a survey and a donation to UNICEF, but does not know the consequences of any specific option. However, the client knows that the advisor had complete payoff information at the time of writing the recommendation. The client is free to choose the recommended option or any other option.

Our conditions vary only the recommendation that the advisor is hired to make. In *neutral work*, the advisor must recommend a specific one of the nine options that is beneficial to the client and to UNICEF. Because a client is very likely to obtain such an outcome independently of any advice, the advice's impact is largely neutral. In *immoral work*, the job is to recommend the single option with negative consequences. In both cases, the advisor makes a recommendation by completing a form stating that the recommended option "will save the highest number of children" and "will give you the highest financial reward." Recommending the harmful option, therefore involves deceit and likely causes the client and UNICEF to lose money.[20] Note that conditions only differ in the moral nature of the job; everything else, including effort costs, is held constant.

The market. Participants are randomly allocated to markets consisting of 6 workers who compete to be hired by 6 automated firms. Each worker can provide up to two units of labor—one at a low cost (CHF 2.50) and one at a high cost (CHF 5.50). All workers face the same induced costs, as clearly explained in the instructions. At the beginning of every market period, each worker decides whether to participate in the market. If participating, the worker then (privately) provides two wage requests, one for each possible labor unit that worker can provide. Workers may only submit wage requests at least as high as the corresponding cost of providing that labor unit.

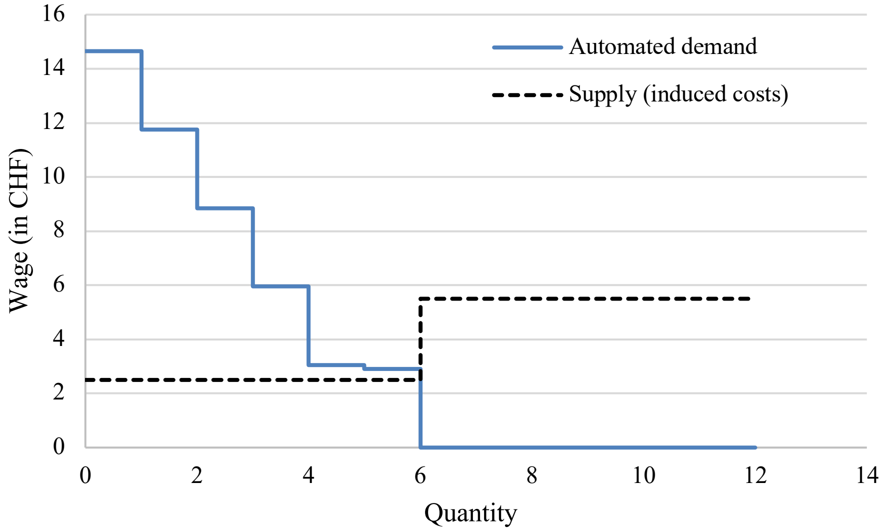
Firms are simulated by the computer. Each firm can hire up to one unit of labor per period. Firms are identical except for the wage that they offer to the workers. Figure 2 displays the automated demand for labor as well as the induced costs of labor supply. In equilibrium, all workers provide one unit of labor and the market wage is between CHF 2.5 and 2.9.[21] The workers have no information about the shape of the automated demand.

[20] We introduced 10 options to create a setting in which "bad" advice is harmful and "good" advice largely inconsequential. A client not receiving a recommendation has a 9/10 chance of selecting a beneficial option, whereas a client who follows bad advice is likely to be harmed. Clients followed the recommendation in 84% of all cases.

[21] We selected this specific labor demand function to facilitate equilibrium convergence. As long as the wage is higher than CHF 3.05, at least two workers will be unemployed, putting downward pressure on wages.

We use a uniform-price sealed-offer auction as the market mechanism, as this provides desirable features. First, Smith et al. (1982) show that this type of market typically converges to the equilibrium prediction. Second, this mechanism allows us to automate labor demand (see also Sausgruber and Tyran, 2011) and therefore hold demand constant across conditions. Once all six workers submit their wage requests, the computer ranks them from lowest to highest and compares the wage requests to the firms' wage offers, ranked from highest to lowest. The *market wage* is the lower of two candidates: (i) the last wage offer higher than the wage request with the same rank and (ii) the first wage request higher than the wage offer with the same rank. This mechanism clears the market—for the market wage, labor supply equals labor demand and all workers with wage requests below the market wage are hired. A worker's earnings in a round equal the market wage times the units of labor provided by that worker (0, 1 or 2), minus the corresponding costs; workers who do not participate in the market earn zero in that period.

Figure 2: The automated demand and the induced costs of labor supply (Study 1)



The market repeats for 15 periods with the same set of workers and in the same treatment condition. After each period, the computer displays the market wage, the picture of every worker in the market, and information regarding each workers' outcomes across all periods (see Appendix Figure A2). This information includes employment outcomes, wages, and cumulative earnings for all workers across periods, connected to the workers' photographs. After observing outcomes, hired workers complete the corresponding recommendations on paper forms—they write their own initials and the letter corresponding to the job requirement.[22] If a firm fails to hire a worker in a period, the firm's client receives no recommendation. This implies that a subject who reduces their labor supply may also (weakly) reduce the number of clients who receive bad recommendations.

5.1.3 Survey study with "clients"

We recruit a separate sample of individuals for two functions. First, they serve as "clients" who act on the recommendations from the laboratory labor market. Each client makes up to six decisions involving choosing one of ten lettered options. They are told that these decisions influence their own earnings and possibly the size of a UNICEF donation. They do not know the mapping from choices to payoffs, which varies across decisions. Clients receive mixtures of recommendations with good advice, bad advice and no advice (from cases in which a firm was unable to hire a worker). Clients only learn the payoff consequences after making all decisions.

Second, these participants also complete a survey in which they rate the (im)morality of various firms and industries. We describe these ratings and how we use them in Section 7.

5.1.4 Procedural details

Instructions and materials are available at https://osf.io/4c6r7/. Our study obtained ethical approval from the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at University of Zurich.

Laboratory sessions took place at the Decision Sciences Laboratory (DeSciL) at the Federal Institute of Technology in Zurich (ETH) in February through May of 2017, using the software zTree (Fischbacher, 2007). We recruited participants using hroot (Bock, Baetge and Nicklisch, 2014) from the joint subject pool of the University of Zurich and the ETH. We only accepted participants who had previously completed the online survey described in Section 7.[23] To start, subjects entered an identifier that allows us to link, anonymously, their answers in the online

[22] Subjects are informed that each firm has a 0.25 probability of having a client in each period, independently of whether the firm hires a worker. If the firm has no client, then the worker's recommendation is unused, although the worker still completes the recommendation and receives the market wage. Subjects do not know whether a firm will have a client in that period at the time they submit wage requests or complete any forms. This represents, for instance, a case in which a worker is hired to prepare promotional materials for a harmful product, which may not be seen by potential customers. At the end of the experiment, subjects learn which of their written recommendations will be distributed. We employ this procedure to reduce the number of clients subsequently needed. It implies that writing a recommendation only has consequences with 0.25 probability, working against our hypothesized treatment effect.

[23] We made an exception if less than 24 subjects who completed the survey showed up to the experiment. In total, three subjects were allowed to participate despite not completing the online survey.

survey to their lab behavior. We conducted ten sessions, resulting in a total of 240 participants, allocated to 28 markets for immoral work and 12 markets for neutral work.

For the role of clients, we recruited a different sample of students (N=177) on the campuses of the University of Zurich and the ETH. We invited passersby to participate in a 5-minute study in which they could earn money (at least CHF 2) and potentially generate a donation for UNICEF.

5.2 Results

We first discuss how we construct our measure of concern for morality, θ. Next, we study behavior and sorting in the labor market, and whether it can be predicted by θ. We then study labor supply and wages in the labor market and their connection to θ.

5.2.1 Construction of θ

We construct θ based on choices in the behavioral task completed at the beginning of the laboratory session (Appendix Table A3 shows the distribution of choices). Let m_{ir} be the number that individual i reports if the actual die roll is r. We classify an individual as a θ_L *type* if $m_{ir} \geq r$ for all $r \in \{1, 2, \ldots, 6\}$ and $m_{ir} > r$ for at least one r; that is, if the participant lies at least once for personal gain and never in a self-harmful manner. We classify the remaining participants as θ_H *types*. Based on this classification, 66 (27.5 percent) are θ_L types and 174 (72.5 percent) θ_H types.[24] For convenience, we often refer to θ_L types as *immoral* and θ_H types as *moral*.

5.2.2 Labor market behavior and sorting

We first test differential sorting into immoral work. Our main prediction is that, assuming θ measures a stable moral concern, we should observe differential labor-market behavior between θ_H and θ_L types, but only when employment requires immoral work. The data confirm this. Table 2 (columns 1 and 2) reports the results of a double-hurdle regression of the decision whether to submit a wage request and, conditionally, the actual wage request. The key independent variable is a subject's type (θ). In the immoral work condition, θ_H types submitted wage requests less

[24] Thirteen subjects (5.4 percent, see Appendix Table A3) harmed themselves at least once with a lie ($m_{ir} < r$). These subjects do not appear to be motivated by egoism, so we classify them as θ_H. The remaining 161 subjects classified as θ_H always report truthfully. Classifying subjects that lied in a self-harmful manner as θ_H types is conservative in that they act less morally than the honest subjects (see Appendix Table A5). Our results do not change if we drop these subjects or if we classify them instead as θ_L. We also find similar results using alternative classification approaches (see Appendix Table A5). For example, we can classify subjects into more than two categories—e.g., conditional on the number of lies or based on the expected payoff from lying ($\frac{1}{6}\sum_{r=1}^{6} m_{ir} - r$). Due to the low number of subjects with different lying-patterns (see Appendix Table A3), we use a binary classification for our main analysis.

frequently, by almost 30 percentage points, than θ_L types (61.6 percent vs. 90.6 percent, p<0.001). By declining to submit a wage request, a subject indicates unwillingness to work even at a wage of 50 CHF per period, the maximum wage request. Furthermore, consistent with the model, θ_L types submit conditional reservation wage requests that are approximately 0.49 CHF lower than those of θ_H types (p=0.073).[25] These effects do not weaken over time.[26] Moreover, 21.5 percent of the θ_H workers never participated (refused to submit wage requests in any period), but this is true of only 4.3 percent of θ_L workers (t = -3.78; p=0.001). Hence, our behavioral measure of subjects' moral types predicts their willingness to seek employment in an immoral job.

Table 2: θ predicts behaviors and outcomes in laboratory labor markets (Study 1)

Dependent variable:	**Participation**	**Reservation wage**	**Employment rate**	**Number of work units**
	(1)	(2)	(3)	(4)
Low-theta (θ_L)	1.024*** (4.64)	-0.494* (-1.80)	0.266*** (5.59)	0.254*** (4.93)
Neutral work (N)		-1.147*** (-5.47)	0.331*** (7.75)	0.272*** (7.15)
θ_L * N		0.476* (1.67)	-0.268*** (-4.57)	-0.256*** (-4.13)
Constant	0.295** (2.41)	4.056*** (20.31)	0.496*** (13.55)	0.555*** (17.88)
N	2,520	2,832	3,600	3,600
Data	Immoral	Immoral + Neutral	Immoral + Neutral	Immoral + Neutral
LL (pseudo)	-1427.9	-6135.8	-	-
R^2	-	-	0.104	0.060
p-value: θ_L + θ_L*N = 0	-	0.822	0.957	0.957

*Notes: Models (1) and (2): Estimates from Craggs double-hurdle Model: (1) probit model; (2) truncated linear regressions (truncated from above at 50 CHF). Note model (1) uses only data from the immoral work condition because non-participation is virtually non-existent for neutral work. Models (3) and (4): Estimates from linear regression models. Independent variables: Low-theta in {0, 1}, Neutral work in {0, 1}. Note that column "p-value: θ_L + θ_L*N = 0" test sorting into neutral work. Standard errors clustered at market level; t-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

In the neutral work condition, however, non-participation is virtually non-existent—there were only 3 total cases, or 0.28 percent of all observations. That is, there is nearly universal participation when the job does not involve immoral behavior. Moreover, the average wage

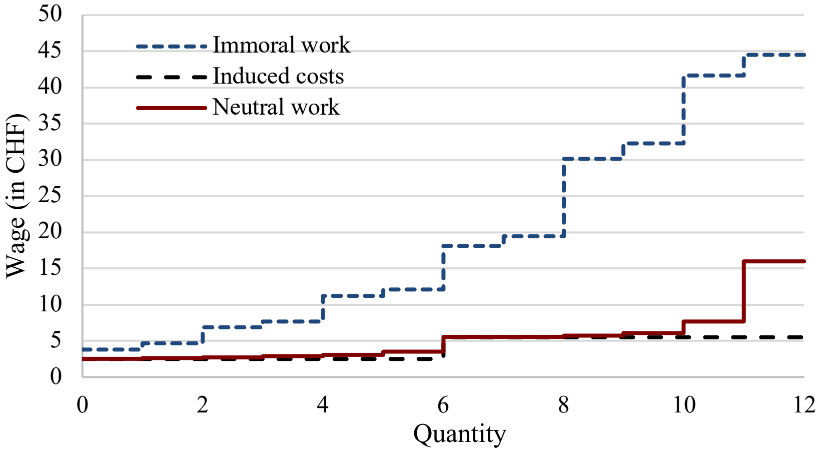
[25] These estimates are from the double-hurdle model. Using raw data, in the immoral (neutral) work condition, average conditional reservation wages are CHF 4.14 (CHF 2.91) for θ_H types and CHF 3.56 (CHF 2.89) for θ_L types.

[26] Specifically, if we add a linear time trend to the hurdle model and its interaction with θ_L (see Appendix Table A4), we find that θ_L types become slightly more likely to participate over time and provide lower reservation wages, relative to θ_H types. However, both coefficients are small and statistically insignificant.

requests of θ_H (CHF 2.91) and θ_L (CHF 2.89) types do not differ in magnitude or in statistical significance (p-value from truncated linear regression = 0.822).

The differences in labor market behavior between θ_H and θ_L types in the immoral work condition also result in different employment rates. Table 2 (column 3) shows a difference of 26.6 percentage points in employment in immoral work between θ_H and θ_L types (p<0.001). The results are similar if we use, as a dependent variable, the number of work units provided (0, 1 or 2) rather than a binary employment measure (column 4).[27] Appendix Figure A3*a* shows that, if anything, the difference in employment rates grows across periods. In the neutral work condition, we find no significant difference in employment rates (columns 3 and 4; Appendix Figure A3*b*). This corroborates that the difference in hiring rates in the immoral work condition is driven by differences in concerns for morality and not some other difference between θ_H and θ_L types.

Figure 3: Labor supply for neutral and immoral work in the laboratory (Study 1)



Notes: Wage requests are ranked within each market period of each group. The figure shows the average wage request for each rank for both the immoral work and the neutral work conditions. Note that wage requests are censored at the maximal wage request that subjects could make, 50 CHF. For this figure, we set the wage requests of subjects who are not willing to participate to CHF 50. Therefore, the supply curve for the immoral work condition should be interpreted as a lower bound.

5.2.3 Labor supply and wage premiums for immoral work

Figure 3 shows that labor supply differs substantially between the two kinds of markets.[28] For neutral work, labor supply is close to the induced costs. However, for any given wage, there is a substantially lower supply of labor for immoral work. These differences in labor supply are the

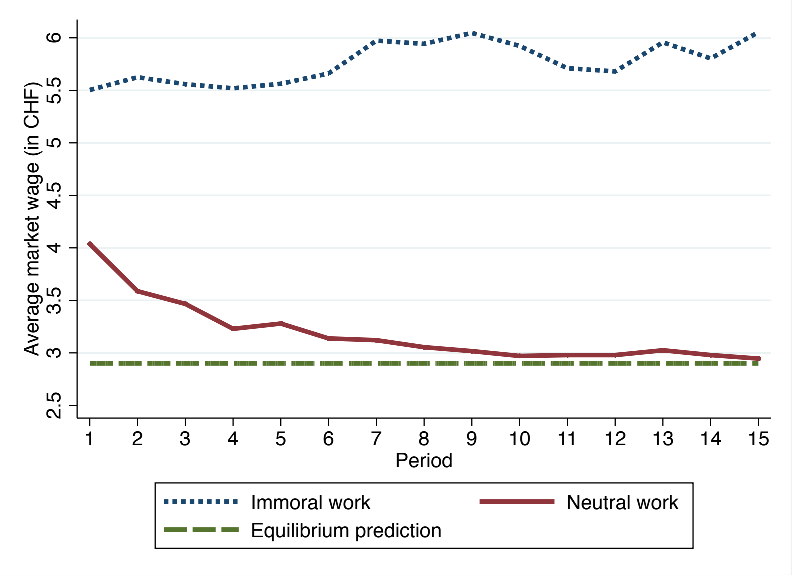
[27] Results are robust to demographic controls and market fixed effects, see Appendix Table A6.

[28] Appendix Figure A4 shows the labor supply if we only consider the last 5 periods. Appendix Figure A5 displays the labor supply in (simulated) labor markets with only low-theta or only high-theta types.

result of the differential labor market behaviors discussed above: when work requires immoral behavior, θ_H types withdraw their participation and make higher wage requests. However, when the work activity is neutral, both types almost always participate and make similar wage requests.

As a direct consequence of the differential labor supply in Figure 3, we find a substantial immorality premium—between roughly CHF 1.5 and CHF 3 per period—as shown in Figure 4. While market wages for neutral work converge toward the equilibrium prediction of CHF 2.90, the average market wage is persistently higher for immoral work and this difference is statistically significant in a t-test from a regression with standard errors clustered at the market-level (coefficient=2.58, t=6.00, p<0.001). Moreover, the difference grows across periods, suggesting that it is robust to repeated market competition and experience. This persistence is remarkable given the information subjects received at the end of each market period: in the immoral work condition, participants who forgo working see other participants repeatedly earning high wages.[29]

Figure 4: Immorality wage premium in laboratory labor markets (Study 1)



6. Study 2: Non-consequentialist preferences in moral and immoral work

Study 1 provides support for our hypotheses of wage premiums and sorting. However, workers can only influence the harm caused by immoral work by reducing their labor supply, thereby

[29] We also find persistent differences in employment levels in the two markets (Appendix Figure A6). While markets for neutral work converge to the equilibrium quantity of 6, the average quantity remains below 4 in the immoral work condition. This difference is significant in a t-test comparing the means (coefficient=-1.201, t=-6.30, p<0.001). Moreover, because of the reduced labor supply by moral types, immoral types are better off in markets for immoral work, particularly when in the presence of more moral subjects (see Schneider, Brun and Weber, 2020, for details).

potentially lowering the number of jobs executed. Study 1 also does not allow individuals hired for immoral work to take actions that reduce the resulting harm from such work. Thus, reduced labor supply by moral workers may result from either a desire to have positive impacts by reducing the quantity of harm produced or an aversion to personally performing immoral work, as proposed by our model. The distinction between these two motives is important for understanding how sorting into immoral work differs from sorting into moral work—for example, whether moral individuals are willing to enter harmful industries if they can reduce the resulting harm, or whether aversion to personally performing immoral work drives out moral individuals from such work.

A more stringent test of our hypotheses is one in which reducing one's labor supply for immoral work does not mitigate harm produced by immoral work but may instead produce greater harm. This is the case, for example, when labor market decisions do not impact the number of people employed in immoral work and when workers hired to perform immoral work can take actions that reduce harm. In this case, workers motivated to reduce overall harm should enter "immoral work" to mitigate harm, possibly even at low wages. This effect could lead the most morally inclined individuals to sort into immoral work and a *negative* immorality wage premium.

Study 2 adds these features. To motivate our investigation, we extend the theoretical framework in Section 3 to allow workers employed in immoral work to lessen the negative impact of the job by an amount $e < I(j)$ at a cost c. We then compare the labor market outcomes for immoral work with those for analogous moral work associated with activities that produce positive impact. For moral work, a worker who is hired can produce positive impact by an amount e at a cost of c. From a consequentialist standpoint, immoral and moral work are equivalent, as both allow a worker to benefit society by an amount e at a cost c. Hence, a model that assumes workers to be motivated only by consequentialist (or, "pure") social motives would predict morally motivated workers to find entering moral and immoral work equally attractive.

The model in Section 3 posits that workers are influenced by an aversion to personally implementing immoral work and a corresponding preference for performing moral work (as with "impure" motives such as warm glow or image concerns). As we show in Appendix C.3, if workers have impure moral motives, the model predicts immorality wage premiums that distinguish markets for immoral work from those for moral work (Proposition 3). The model also predicts differential sorting for the two types of work—i.e., immoral types sort into immoral work and moral types sort into moral work (Proposition 4).[30] As a result of such sorting, the model also predicts that the costly moral action at work will be implemented more frequently in moral work than in immoral work. Study 2 tests these predictions in a laboratory experiment. Note that our model only makes qualitative predictions for labor markets for moral and immoral work, and does not specify relationships between magnitudes of wage or sorting effects across such markets.

6.1 Design

The design of Study 2 follows that of Study 1. Appendix Figure A1 gives an overview of the sequence of the study, and a comparison to the sequence of Study 1. We first measure participants' concern for morality, θ, using a behavioral task. Subsequently, subjects participate in a repeated laboratory labor market, which operates similarly to the labor market for Study 1.

We implement several key changes relative to Study 1. First, we fix the market quantity at four units, using an inelastic demand for four jobs and "computer workers" who substitute human subjects' labor supply above a wage threshold. Thus, all four jobs are filled even if all participants refuse to provide immoral work. Second, in addition to immoral and neutral work, Study 2 introduces a third treatment with moral work. Third, workers in the immoral work condition can reduce the harm produced by their work at a personal cost, while workers in moral work can produce a comparable increase in the positive impact of their work at the same cost.

We also change two additional design features. First, to create a closer connection to the kinds of real-world jobs that people perceive as immoral, we use a context reflecting working to impact the availability of guns—similarly, for example, to working in the weapons industry. Specifically, hired workers in the immoral work condition perform two activities. First, they send messages that attempt to minimize, through misinformation, the harm produced by gun violence. Second, they produce donations for the National Rifle Association (NRA)—a non-profit advocating for reduced firearms regulations—at the expense of donations to Everytown for Gun Safety (Everytown)—a non-profit advocating for gun safety.[31] Second, in all three treatments we

[30] One may be concerned that moral subjects try to maximize their lab income to make a larger donation to charity outside the laboratory. Such a motivation is unlikely to be a problem for our study, as it would work against finding evidence of wage premiums and differential sorting based on moral concerns. Moral workers wishing to generate high earnings to then donate outside the lab, should do so in both the conditions involving moral and immoral work, and, in fact, more so in the latter whenever there is a wage premium, as is the case in both our studies.

[31] By using donations to the NRA to investigate immoral behavior, we follow Ariely, Bracha and Meier (2009). Note that we designed the experiment so that although subjects transferred substantial donations between the two organizations, the final donations to the NRA and Everytown were balanced and, more importantly, small and largely inconsequential. The final donations, paid privately be the researchers, were CHF 218 (264) to the NRA (Everytown).

introduce a second, morally neutral, job—all workers not hired in the primary labor market of interest complete a morally neutral job at fixed wages. This incorporates a realistic feature of labor markets and mirrors the theoretical model used to guide Study 2 (see Appendix C.3).

In the following, we provide additional details for some of the key features of Study 2. We provide a more detailed discussion of the design of Study 2 in Appendix D.

6.1.1 Behavioral measure of concern for morality (θ)

As in Study 1, participants first play an incentivized game that measures their willingness to lie for personal gain while causing harm to others in a non-market environment. We use the same game as in Study 1 but adapt the harmful impact to potentially affect donations to the NRA and to Everytown. Specifically, if Participant A sends a dishonest message and Participant B chooses to trust this message, then Participant B earns less and a donation to Everytown is transferred to the NRA. In contrast with Study 1, we change the Participant B to a participant in a subsequent study.

6.1.2 Labor market experiment

In the labor market, subjects play the role of workers competing to be hired by automated firms.

The jobs. There are two types of jobs in each round, *neutral jobs* and *G jobs ("Gun jobs")*. Neutral jobs require a simple task (moving a LEGO brick), which has no consequences. We vary the moral nature of the G job by treatment—with *immoral*, *moral*, morally *neutral* work conditions. In each condition, the G job always consists of two elements: selecting a statement to disseminate and a reallocation task. We next discuss these two elements of the G job.

Selecting a statement: In both the immoral and moral work conditions, participants are tasked with disseminating (mis-)information related to gun violence. Those employed for a G job select one of two statements displayed on the computer screen—one asserting that gun violence is the leading cause of death for minors in the U.S. and the other claiming that gun violence is an uncommon cause of such deaths. Participants are informed that only the former statement is accurate (Goldstick, Cunningham, & Carter, 2022). In the immoral (moral) work conditions, workers must select the false (true) statement. Subsequently, the chosen statement is presented to U.S. participants in a separate online survey study.[32] We added this feature to increase "work

[32] We informed lab participants that we would subsequently recruit US residents to participate in an online survey. These survey participants would be told that they would see a statement selected to be shown to them by a participant in an earlier study and would subsequently state their attitudes toward gun laws. We informed lab participants (truthfully) that we would not reveal the true statement to the US participants. It is important to note, however, that

immorality" from the perspective of a non-consequentialist. However, as we discuss below, the number of G jobs implemented—and therefore also the number and content of statements sent—is independent of workers' labor market decisions. This job feature is thus irrelevant to a consequentialist but relevant to non-consequentialists. In the neutral work condition, workers hired for a G job must select a correct statement from two morally neutral options.

The reallocation task: In the moral and immoral work conditions, workers hired to perform the G job also reallocate donations between the NRA and Everytown, with discretion over the size of the reallocation. In immoral work, they replace existing donations earmarked for Everytown with donations to the NRA.[33] For each job a hired worker performs, the worker chooses whether to reallocate CHF 0.6 from Everytown to the NRA (a net change of CHF 1.2) or to reallocate CHF 0.4 (a net change of CHF 0.8). Workers can thus impact the size of donations for the NRA and for Everytown. However, the option that produces less negative social impact is costly—a worker incurs a cost of CHF 0.2 for reallocating the smaller amount. In the moral work condition, the reallocation works in the opposite direction, moving either CHF 0.4 or CHF 0.6 from the NRA to Everytown, with the larger reallocation incurring a CHF 0.2 cost for the worker. To provide a tangible component to the work, we operationalize the reallocation task by having workers replace LEGO bricks of one color with those of a different color (see Appendix D for details). In the neutral work condition, workers hired to the G Job reallocate LEGO bricks with no consequences.

The market. Participants are randomly allocated to markets consisting of 6 workers and 4 "computer workers" who can compete to be hired for four "G jobs" or can work in outside neutral work ("N jobs"). Each worker provides two units of labor. The wages for doing the neutral jobs are fixed: workers earn CHF 1.1 for doing one neutral job and CHF 0.50 for a second neutral job.

The market repeats for 12 periods.[34] In a market period, each worker decides whether to participate in the labor market for G jobs and, if so, provides a wage request for each of the two possible labor units. The three computer workers each always submit reservation wages of CHF

we clearly informed the US survey participants that the observed statement had been selected by another participant and that it could be true or false. We also provide a link at the end of the online survey to access the truthful underlying information. Our study was reviewed and approved by the Human Subjects Committee at the University of Zurich.

[33] The NRA is very unpopular with our participants and transferring donations to the NRA is generally viewed as "immoral." We asked participants how ethical or unethical they think it is to take actions in the experiment that increase a donation to the NRA and decrease a donation to Everytown. Only 11.6% view such actions as ethical.

[34] The results of Study 1 show that markets converged after 12 periods. We thus reduced the number of periods to 12 to compensate for the added duration of Study 2 due to the added work activity.

15 for up to two G jobs, meaning that any wage requests by a human worker above this amount guarantee that the worker will not be hired.

Labor demand is static across periods, is simulated by the computer, and is completely inelastic. Specifically, "firms" hire exactly four workers to perform G jobs at the market wage. As in Study 1, we use a uniform-price sealed-offer auction as the market mechanism. The equilibrium prediction under standard selfish preferences is that four different human workers are hired to provide one G job each and the market wage is CHF 0.5.

At the end of each period, the computer reports the market wage and the total donation amount replaced by the workers hired to implement four G jobs. As in Study 1, the report displays every worker's picture and summarizes information regarding each workers' outcomes, including wage earnings and employment, across all periods.[35] Workers then implement any jobs for which they have been hired. A computer worker hired for a G job does not take the costly action to reduce harm (increase the benefit) in the immoral (moral) work condition. Workers are fully informed about the behavior of computer workers (reservation wages and moral conduct at work).

6.1.3 Procedural details

All sessions took place at the Laboratory for Experimental and Behavioral Economics at the University of Zurich in October and November of 2023.[36] We recruited participants using hroot (Bock, Baetge and Nicklisch, 2014) from the joint subject pool of the University of Zurich and the ETH. Sessions consisted of 24 participants.[37] We conducted fifteen sessions, resulting in a total of

[35] Workers cannot see the donations reallocated by any individual worker. To prevent disclosing this information through individual earnings, we display only workers' wage earnings (before subtracting potential costs based on the reallocation decision). This corresponds to a situation in which where an individual is employed is observable, but not the precise actions that individual takes at work. Additionally, in our model, a worker's "social image" is tied to the industry or employer's image, $I(j)$, which is best represented by the total amount of donations reallocated.

[36] We also conducted a pilot study in September 2023, consisting of one immoral and one moral session each. This pilot differed in some details, notably in the reporting screen at the end of each period. Despite the low statistical power (N=48), we find evidence for an immorality wage premium and sorting. In the immoral work condition, immoral types were hired 9 percentage points more frequently for G jobs compared to the moral work condition, and the average market wage was CHF 0.47 higher in the immoral work condition than in the moral work condition. Adding these pilot data to the main analysis does not change any of the results.

[37] One session in the moral work condition only had 18 participants (3 markets). In one of the immoral work sessions, a participant chose to opt out of the experiment due to moral concerns. When the participant privately communicated these concerns to us, we invited the participant to quietly step outside the room. Despite being fully aware that this decision would not impact the final donations to the NRA or Everytown, the participant preferred to forgo substantial earnings from the experiment rather than participate. Note that this behavior aligns closely with the type of conduct we theoretically model and investigate in Study 2. We replaced this participant with a research assistant who was instructed to select not to participate in the market for G jobs in each round—consistent with the participant's preferred choice had we enforced participation. We retain this observation in our data to prevent selection bias. However, none of our results change when excluding this participant or all periods of the corresponding market or session.

354 participants, allocated to 24 markets for immoral work, 23 markets for moral work and 12 markets for neutral work. We implemented the experiment with zTree (Fischbacher, 2007). Our study obtained ethical approval from the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at University of Zurich.

Participants received all instructions both on paper and with pre-recorded audio files. We preregistered the data collection, hypotheses, and analysis. Pre-registration, instructions and materials are available at https://osf.io/4c6r7/.

6.2 Results

Our pre-registered analysis focuses on testing our two main hypotheses, the existence of an immorality wage premium and sorting by heterogeneous moral types, as well as and additional hypothesis on moral conduct at work. First, we construct a measure of concern for morality based on the behavioral task from the beginning of the session, using the same approach as in Study 1 (see Section 5.2.1; Appendix Table A7 shows the distribution of choices). Our sample comprises 190 individuals (53.7 percent) classified as having low moral concerns (θ_L) and 164 (46.3 percent) as having high moral concerns (θ_H).

6.2.1 Labor market behavior and sorting

We first study market behaviors and sorting by different moral types. Columns (1) and (2) in Table 3 report the results of a double-hurdle regression of the decision whether to submit wage requests for G jobs and, conditionally, the actual wage request.

The results support the model's predictions. In the immoral work condition, θ_H types submitted wage requests for G jobs less frequently than θ_L types (79.4 percent vs. 96.8 percent, p-value from probit regression = 0.001) and, conditional on submitting wage requests, these requests are 0.85 CHF higher than those of θ_L types (CHF 1.98 vs CHF 1.13, p-value from truncated linear regression = 0.007). These effects do not become weaker over time; while reservation wages fall in the first few periods, they do so to the same extent for θ_L and θ_H types.[38]

[38] In a model with a linear time trend and its interaction with θ_L the interaction is small and statistically insignificant (Probit: coefficient = -0.006, z=-0.24, p=0.808; Truncated linear regression: coefficient=0.046, z=0.72, p=0.470). However, the time trend is statistically significantly negative (Truncated linear regression: coefficient=-0.140, z=-2.33, p=0.020), indicating that reservation wages decrease over periods similarly for θ_L and θ_H types. This decrease occurs in early periods; restricting data to the last five periods yields small and statistically insignificant time trends.

In both the moral and neutral work conditions, we find no differences in market behaviors between moral and immoral types. Both θ_H and θ_L types are equally likely to participate (moral: 96.3 vs. 97.8 percent, p-value from probit regression: 0.345; neutral: 98.0 vs. 97.1 percent, p-value: 0.614) and they submit similar wage requests on average (moral: CHF 0.81 vs. CHF 0.76, p-value from truncated linear regression: 0.575; neutral: CHF 1.28 vs. CHF 1.13, p-value: 0.697).[39]

Table 3: θ predicts behaviors and outcomes in the laboratory labor markets (Study 2)

Dependent variable:	**Participation**	**Reservation wage**	**Employment rate**	**Number of work units**
	(1)	(2)	(3)	(4)
Low-theta ($\boldsymbol{\theta_L}$)	1.030***	-0.778***	0.215***	0.189**
	(4.36)	(-2.69)	(4.18)	(2.48)
Moral work (M)	0.959***	-1.099***	0.156***	0.086*
	(4.41)	(-4.09)	(3.93)	(1.80)
Neutral work (N)	1.238***	-0.626*	0.152***	0.078
	(5.32)	(-1.93)	(3.83)	(1.59)
$\boldsymbol{\theta_L}$ * M	-0.805**	0.725**	-0.168***	-0.159*
	(-2.40)	(2.39)	(-2.67)	(-1.85)
$\boldsymbol{\theta_L}$ * N	-1.200***	0.633	-0.140**	-0.138
	(-2.92)	(1.34)	(-2.00)	(-1.47)
Constant	0.821***	1.907***	0.438***	0.564***
	(5.26)	(7.41)	(13.27)	(13.51)
N	4,248	3,985	4,248	4,248
LL (pseudo)	-849.4	-8928.3	-	-
$\mathbf{R^2}$	-	-	0.026	0.011
p-value: $\boldsymbol{\theta_L + \theta_L}$*M = 0	0.345	0.575	0.208	0.459
p-value: $\boldsymbol{\theta_L + \theta_L}$*N = 0	0.614	0.697	0.125	0.352
p-value: M = N	0.225	0.025	0.885	0.829
p-value: $\boldsymbol{\theta_L}$*M = $\boldsymbol{\theta_L}$*N	0.338	0.811	0.647	0.750

*Notes: Models (1) and (2): Estimates from Craggs double-hurdle Model: (1) probit model; (2) truncated linear regressions (truncated from above at 50 CHF). Models (3) and (4): Estimates from linear regression models. Specification (3) is the pre-registered main specification. Independent variables: Low-theta in {0, 1}, Moral work in {0, 1}, Neutral work in {0, 1}. Note that columns "p-value: θ_L + θ_L*M = 0" and "p-value: θ_L + θ_L*N = 0" test sorting into moral and neutral work, respectively. Columns "p-value: M = N" and "p-value: θ_L*M = θ_L*N" test treatment differences in sorting between the moral and neutral work conditions. Standard errors clustered at market level; t-statistics in parentheses; * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.*

Next, we study whether differences in labor market behaviors translate into differences in employment. In Table 3, column 3, we present results from a linear regression of employment rates for G jobs on θ, our pre-registered main analysis for studying sorting. On average, θ_L types are 21.5 percentage points more likely to be employed than θ_H types for immoral work (p<0.001).

[39] Table 3 also compares sorting across conditions. The differences in market behavior between θ_L and θ_H types in the immoral condition (captured by the coefficient for θ_L) are statistically significantly different from the corresponding differences in the neutral and the moral conditions (captured, respectively, by θ_L * N and θ_L * M).

In the moral and neutral work conditions, there is no significant difference in employment rates between the two types (p-values = 0.208 and 0.125, respectively). Importantly, we can reject that the differences in hiring rates between the types ("sorting") are the same in the immoral and the moral work conditions (p = 0.010) and in the immoral and the neutral work conditions (p = 0.050). We obtain qualitatively similar results when we use, as a dependent variable, the number of G jobs provided (Table 3, column 4), when we employ alternative methods to categorize individuals into moral and immoral types (Appendix Table A8), and when add demographic controls and market fixed effects (Appendix Table A9). Appendix Figure A7 illustrates that the difference in employment rates between θ_H and θ_L types in the immoral work condition persists across rounds.

The study thus supports the model's first main prediction that immoral types sort into immoral work, and that such sorting differs from sorting into moral work. Consistent with the non-consequentialist moral preferences in the model, moral types avoid immoral work even in settings where labor market behaviors do not impact the quantity of immoral work performed and when employment provides an opportunity to reduce the harm resulting from immoral work. Interestingly, we find that moral types do not sort into moral work.

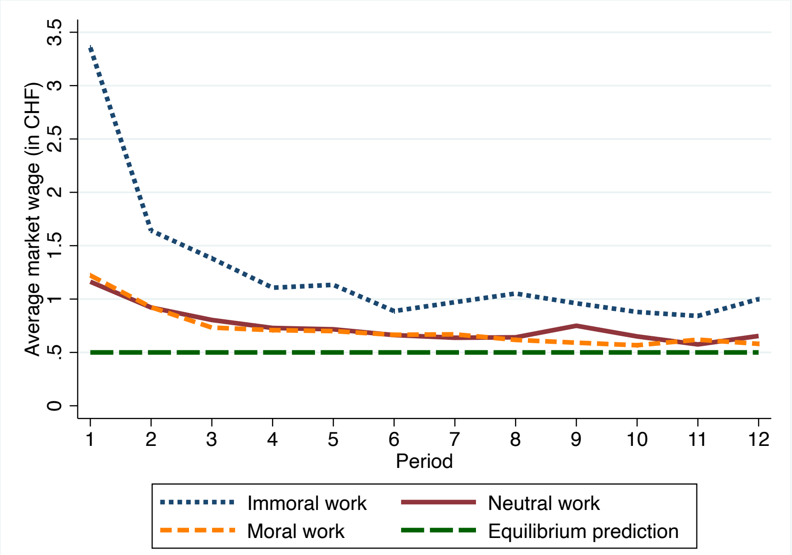
6.2.2 Wage premiums for immoral work

Our second focus is on whether we observe wage premiums for immoral work. Figure 5 reveals that market wages in the moral and neutral work conditions approach the equilibrium prediction of CHF 0.50, but that immoral work yields an immorality wage premium of approximately CHF 0.50. Market wages in the immoral work condition are statistically significantly different from both the wages in the moral work condition (OLS regression with standard errors clustered at the market-level: coefficient=0.55, t=3.57, p=0.001) and the neutral work condition (coefficient=0.53, t=3.33, p=0.002). Market wages decrease initially in all three treatment conditions, consistent with the initial decline in reservation wages, but stabilize after a few rounds. In the last 5 rounds, market wages in the immoral work condition average CHF 0.95, which is almost twice the equilibrium prediction, and 1.59 times the average wage in the moral work condition in the last 5 periods.[40]

[40] If we regress market wages on a linear time trend (standard errors clustered at the market-level) wages are estimated to decrease by CHF 0.13 per round in the immoral work condition (t= -3.48, p=0.002), by CHF 0.04 in the moral work condition (t=-5.15, p<0.001), and by CHF 0.03 in the neutral work condition (t=-3.65, p=0.004). For the last 5 periods, the linear time trends are small and statistically insignificant. Focusing only on the last five periods, we find statistically significant differences in wages between the immoral and moral work conditions (coefficient=0.351, t=2.66, p=0.011) and between the immoral and neutral work conditions (coefficient=0.293, t=2.15, p=0.038).

Interestingly, there are no statistically significant differences in market wages between the moral and neutral work conditions (coefficient=-0.03, t=-0.51, p=0.616). This mirrors the participants' similar market behaviors in the moral and neutral work conditions (see Table 3). Consequently, our findings indicate that while the reluctance to personally implement immoral work is sufficiently strong to influence market behaviors and outcomes, the same does not hold for the preference to personally implement moral work.

Figure 5: Immorality wage premium in laboratory labor markets (Study 2)



6.2.3 Moral conduct at work

Finally, we investigate the "moral action" in the immoral and moral work conditions that reduces the net NRA donation by CHF 0.4 at a personal cost of CHF 0.2. Our model predicts the moral action to be chosen less frequently in the immoral work condition compared to moral work, due to differential sorting into moral and immoral work and the assumptions that the cost of the moral action, c, is the same in both conditions and that immoral types are less likely to choose the moral action at work. In line with the later assumption, θ_L types indeed select the moral action less frequently than θ_H types (15.96% vs. 40.0% of all cases; t=-4.53, p<0.001).[41]

Surprisingly, despite the overall positive relationship between moral concern and moral

[41] Results are from linear regressions with standard errors clustered at the market-level. We pool the data from the immoral and moral work conditions. We also run a regression where we add an immoral work condition treatment dummy, and the interaction between the treatment dummy and θ_L; the interaction is not statistically significantly different from zero (coefficient = -0.147; t=-1.61, p=0.115).

behavior at work, Table 4 shows that hired workers choose the moral action 17.2 percentage points more frequently when hired for immoral work than moral work (p=0.008). Appendix Figure A8 shows that this pattern persists across rounds. The data thus contradict the model's third prediction.

We cautiously provide three potential *post hoc* interpretations for this unexpected finding. First, as we note earlier, market wages for G jobs are substantially lower in the moral work condition than in the immoral work condition, raising the possibility of an income effect, whereby higher wages for immoral work facilitate incurring the cost of the moral action. This is consistent with a few studies finding morality to be a normal good (e.g., Andreoni, Nikiforakis and Stoop, 2021; Bartling, Valero and Weber, 2021). The second and third potential mechanisms are moral licensing and moral cleansing effects (e.g., Merritt, Effron and Monin, 2010; Tetlock, et al., 2000). Participants in the immoral work condition accepting "immoral" jobs might view the moral action at work as a relatively inexpensive means to repair their perception of their own morality. Conversely, participants in the moral work conditions hired for "moral" jobs might feel licensed to behave selfishly at work.

While our experimental design does not allow us to study moral licensing and moral cleansing effects, we provide some evidence for the potential role of income effects. In Table 4, column (2), we study the treatment differences in moral behavior when controlling for market wages. The results indicate that higher market wages are indeed associated with a higher probability of selecting the moral action.[42] Moreover, accounting for treatment differences in market wages substantially reduces differences in the moral action. The remaining effect, while not statistically significant, may be driven by moral cleansing and licensing effects.

In the model, these income, moral cleansing and moral licensing effects can be captured through the net costs of the moral action, c. Although we equalize the monetary costs of the moral action in our experiment, our findings raise the possibility that psychological costs may vary based on factors that differ between moral and immoral work, such as wage income and moral cleansing and licensing opportunities. Appendix C.3 presents a model in which c can vary between immoral and moral work and differences in such psychological costs can dominate effects from differential

[42] We use the logarithm to compress very high market wages in early market rounds. Results are robust to different ways of controlling for market wages. We can, for example, use a more flexible functional form of wages by controlling for wage deciles. Then, the treatment effect on the frequency of the moral action reduces to 0.098 (t=1.55, p = 0.127). We also run a specification in which we control for a dummy that is 1 if the market wage is below 15 and 0 otherwise. Then, the treatment effect is 0.119 (t=1.75, p=0.088).

sorting into immoral and moral work.[43] The generalized model captures the findings of study 2: an immorality wage premium, differential sorting, and more moral conduct at immoral work.

Of course, one must be cautious when considering post hoc interpretations for unexpected results. Rather than definitive interpretations, we present the above as possibilities for why moral behavior at work may be more frequent for immoral work that involves overall negative impacts.

Table 4: Moral behavior at work (Study 2)

Dependent variable:	**Chose costly moral action at work**	
	(1)	(2)
Immoral work	0.172*** (2.75)	0.099 (1.65)
ln(market wage)		0.182*** (3.69)
Constant	0.169*** (4.41)	0.241*** (5.07)
N	2,256	2,256
R^2	0.039	0.074

*Notes: Coefficient estimates of linear regression models. Independent variables: Immoral work in {0, 1}, ln(market wage) is the natural logarithm of the market wage for G-jobs. Standard errors clustered at market level; t-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

7. Stated real-world employment preferences and sorting

In Studies 1 and 2, individuals with the lowest concerns for morality sort into immoral work in a laboratory labor market. In this section, we present complementary evidence of similar sorting for labor markets outside the laboratory. We use data from an online survey conducted as part of Study 1, completed by participants several days (4-7) before the laboratory session. This survey includes questions measuring subjects' expectations of their own future labor market outcomes, including the willingness to work for different firms and industries and expected future wages.[44]

[43] An alternative explanation is that workers hired for G jobs differ across conditions in unobservable characteristics. Our data allows us to discriminate between such an explanation and explanations, like those above, in which an individual hired to do a G job behaves differently in the moral and immoral work conditions. The former explanation would imply that, conditional on being hired, an individual is equally likely to select the moral action in both conditions. We calculate, for each participant, the relative frequency with which the participant selected the moral action when hired to do a G job. The income/licensing/cleansing explanations imply that the distribution of this variable differs between the moral and immoral conditions, while the explanation based on differences in unobservable characteristics would imply no differences in the distributions of the entire worker populations across conditions. Appendix Figure A9 shows that there are substantial treatment differences in the relative individual-level frequencies of the moral action when considering the entire population of potential workers, suggesting that our findings are driven by a change in behavior induced by the different market conditions, rather than purely through differential sorting.

[44] The online survey also includes psychological survey items measuring subjects' concern for morality, giving us a second, easily scalable, measure of moral concern. In Appendix E, we discuss the construction of this second measure

In the following, we first discuss how we measure job preferences in labor markets outside the laboratory and then show that θ, as measured with the behavioral task in Study 1, predicts stated job preferences, consistent with the predictions regarding sorting (Propositions 2 and 4).

7.1 The online questionnaire

Subjects answered several questions about their future labor-market expectations. They saw a list of 26 well-known companies in Switzerland and another list of the 20 industries in Figure 1*a*. The lists are available in Appendix Tables B1 and A10. Subjects rated their willingness to work for each firm and industry (1: *not at all willing;* 5: *very much willing*). These were the first questions participants answered, meaning that when they encountered them, they had not seen any references to morality or moral behavior. Subjects next encountered several multi-item scales to measure concern for morality and moral acts. Thorough descriptions of these survey scales are provided in Appendix E. We implemented the online questionnaire with the Qualtrics software.

7.2 Does $\boldsymbol{\theta}$ predict stated real-world labor market preferences?

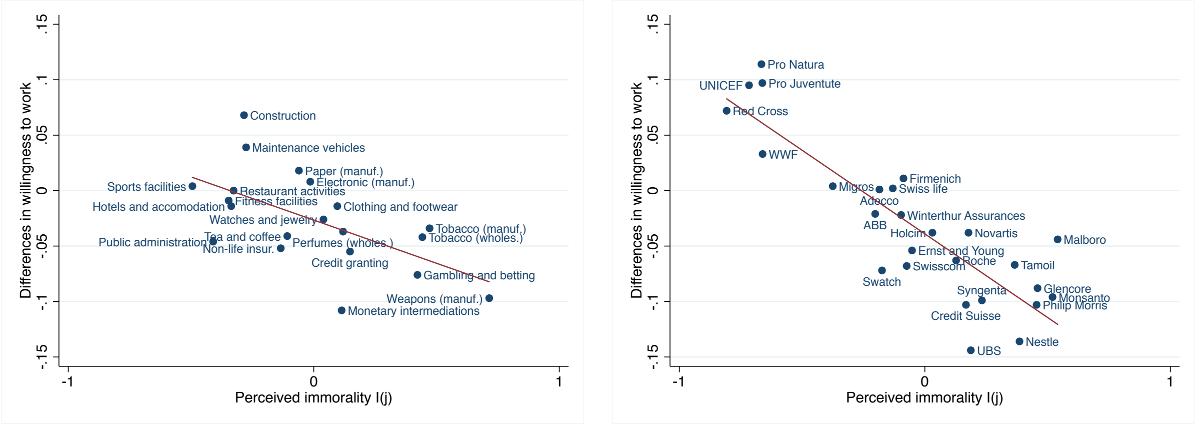
We also separately obtained independent ratings of the perceived immorality of the above firms and industries (see Section 5.1.4). We create measures of *perceived firm immorality* and *perceived industry immorality* by scaling the ratings such that they lie between -1 (very moral) and +1 (very immoral), and then averaging them. We use these as noisy measures of the immorality of work, $I(j)$, in industry (or, firm) j, a key component of our theoretical analysis. The horizontal axis in Figure 6*a* plots the resulting normalized ratings for industries, the horizontal axis in Figure 6*b* plots the normalized ratings for firms (see also Appendix Tables B1 and A10).

Our focus in this section is to investigate whether these perceptions of industry and firm immorality interact with our subjects' concern for morality (θ) to produce differential labor market preferences. For this purpose, we rescale subjects' stated willingness to work for firms and industries, such that they take values between 0 (*not at all willing*) and 1 (*very much willing*).[45]

of concern for morality and show that it correlates with the comparable behavioral measure from the laboratory experiment (θ), with outcomes in the laboratory labor market of Study 1, and with real-world employment preferences.

[45] The industry list accidentally omitted five industries for five participants in the first lab session, meaning we are missing these data. Other than these cases, all subjects completed the full questionnaire. We exclude these missing observations from the analysis. When clients rated the perceived immorality of *firms* and when workers rated their willingness to work for *firms*, they could also choose "I don't know this organization." Our results are similar if we omit responses by clients unfamiliar with a firm when constructing $I(j)$, when we classify such worker observations as "indifferent" or if we restrict our analysis to workers who know all firms (see Appendix Tables A12 and A13).

Figure 6: Correlation between the difference in willingness to work between moral and immoral types and perceived immorality of industries/firms (Study 1)



(a) Industries *(b) Firms*

Source: Survey study (Perceived immorality), online survey (Willingness to work), laboratory experiment (θ^{Exp})
Notes: Differences in willingness to work: Coefficient estimates of linear regression models of the participants' willingness to work for different industries (a) or firms (b) on θ. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: we use θ to classify participants, where θ_H=0 for low- theta types and θ_H=1 for high-theta types. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral.

The vertical axis of Figure 6*a* plots the difference in willingness to work in an industry between subjects classified as moral or immoral, according to θ, while the horizontal axis contains the perceived industry immorality ($I(j)$, obtained from a separate group of respondents). Figure 6*b* shows the corresponding relationship for willingness to work in firms classified as moral or immoral. The strong negative relationships in both figures indicate that subjects classified as immoral are, on average, more willing to work for industries and firms that others perceive as immoral. Appendix Table A11 provides statistical evidence supporting the relationships in Figure 6. We regress a subject's willingness to work for an industry or firm on the perceived industry or firm immorality ($I(j)$), the subject's concern for acting morally (θ) and the interaction of these two terms. While there is little evidence of a systematic difference in willingness to work for neutral industries or firms between moral and immoral types, subjects' moral types strongly predict their willingness to work in industries or firms perceived as immoral. This pattern is significant at least at the 5%-level, and is robust to the inclusion of sociodemographic and industry controls.

The above analysis provides evidence, from outside the laboratory, supporting our second main prediction. Those who are least concerned with morality are significantly more willing to work in firms generally perceived as less moral. While this analysis is based on hypothetical future choices, Wiswall and Zafar (2018) and Gill, Heinz, Schumacher and Sutter (2020) provide evidence that such stated preferences are predictive of ultimate employment. To further validate subjects' stated real-world labor market preferences, we test whether stated employment preferences correlate with employment rates in the immoral work condition in Study 1. Indeed, subjects hired more often for immoral work in the lab have a statistically significantly higher stated willingness to work in immoral industries outside the laboratory (see Appendix Table A14).

8. Discussion and Conclusion

We investigate how the perception that certain types of occupations are immoral and heterogeneity in aversion to personally implementing "immoral" work interact to affect labor market outcomes. We study two key predictions that arise from simple theoretical analysis—first, that jobs generally perceived as immoral yield a wage premium and, second, that individuals less concerned with acting morally sort into such jobs. To test these hypotheses, we employ laboratory experiments, surveys and administrative data in which we measure individuals' heterogeneous concerns for moral conduct and create (or measure) variation in the perceived immorality of different jobs.

Two laboratory experiments provide support for both our hypotheses in environments that vary only the degree to which work involves characteristics generally associated with immoral conduct—social harm and dishonesty. In both studies, markets for immoral work yield higher wages and higher employment rates for workers that we classify as "immoral" based on separate behavioral tasks, a relationship that disappears in labor markets for neutral work or for moral work.

Our second laboratory experiment additionally sheds light on the motives underlying workers' aversion to immoral work. In Study 2, workers cannot reduce overall harm by forgoing employment and can mitigate some harm by seeking employment in "immoral" work. In terms of consequentialist considerations, immoral work is identical to "moral" work in the opportunity for positive impact; but non-consequentialist motives, such as image concerns or warm glow, may nevertheless lead workers to find such work aversive. Our finding of wage premiums for immoral work and sorting out of employment by moral types in Study 2 suggests non-consequentialist motives to be more important. This also means that the sorting and wage patterns in the "immoral" labor market are distinct from those in markets for "moral" work, consistent with the predictions of our theoretical analysis in which non-consequentialist motives drive labor market choices.

Our laboratory experiments provide clear causal evidence that variation in the degree to which work involves acts generally perceived as immoral produces our hypothesized wage and sorting patterns. To address the important question of generalizability from the lab to the field, we also investigate the relevance of our two main hypotheses for non-laboratory labor markets. First, we separately use survey responses, including from a representative sample of the Swiss population, to classify the perceived immorality of real-world industries and show that industries classified as immoral tend to pay higher wages. Second, we also show that those participants in our first experiment whose behavior leads us to classify as immoral tend to express more positive preferences toward working in firms and industries that others generally perceive as immoral.

Our approach of combining varied data sources represents a valuable way of documenting the economic significance of phenomena, establishing causal identification and exploring mechanisms. While laboratory researchers too often do not explore whether their findings are relevant for settings outside the laboratory, researchers using observational data often do not consider using laboratory experiments as a powerful method for generating complementary data.

Our work has potentially important implications. First, in industries with the potential to create societal harm, social welfare may depend on workers' voluntarily internalizing the negative impacts of their actions and forgoing potentially profitable opportunities. For instance, a weapons manufacturer may restrict sales to conflict areas if doing so is sufficiently aversive to top management. One potential implication of our finding that the *least* moral types sort into these industries is that it may be less likely that such immoral types internalize the harm of their actions. As a secondary hypothesis for our second study, we measure the degree to which markets for "moral" and "immoral" work end up producing differential societal harm. Surprisingly, we find that those workers who enter immoral work end up taking more costly moral actions once employed than those who sort into moral work. We cautiously provide a potential interpretation based on income, moral licensing and moral cleansing effects, which can be modeled as the psychological cost of taking moral actions decreasing when sorting into doing well-paid immoral work. Of course, this finding and our interpretation require further study, though they support our broad conclusion that there is something different between the sorting and behavioral patterns that arise in markets associated with immorality, relative to those yielding positive moral impacts.

A second implication of our empirical findings is that the perception that certain firms, industries or types of work are immoral may have secondary impacts on employers, such as how much they have to pay workers and the types of workers they attract. Specifically, our two main results suggest that firms perceived as immoral will pay higher wages and will disproportionately attract workers inclined to care less about moral conduct. Going further, even if particular firms or industries do not engage in anything inherently immoral, the mere perception that this is the case may have consequences. Hence, our findings suggest that managing external perceptions—e.g., "greenwashing"—may be important beyond its impact on the firms' consumers. This provides one reason why firms, like those in the tobacco industry, engage corporate image campaigns targeting worker recruitment (British American Tobacco, 1998, p. 2; Philip Morris International Inc., 1999).

Our theoretical analysis predicts—in line with our experimental data—that the least moral types are overcompensated by the immorality premium. This is in stark contrast to Mankiw's (2010) "just deserts theory"—that is, everybody should receive his or her contribution to society. Our findings suggest a perverse case in which those willing to do the most socially harmful acts may instead benefit from doing so. Moreover, this benefit is the direct result of the actions by others who are concerned with behaving morally and shun employment in work perceived as immoral. At the same time, however, those firms with reputations for negative social impacts may also face higher labor costs, a form of partial internalization.

Of course, our work leaves open many important questions. For example, we do not investigate the precise characteristics that lead some work to be perceived as immoral. While we shed light on the nature of the motives underlying selection into immoral work, more needs to be done to better understand the specific nature of the moral preferences underlying workers' market behavior and how such preferences influence the actions they take as employees in moral and immoral industries. Indeed, our surprising finding in Study 2 that workers who appear relatively unconcerned with moral behavior in one task disproportionately sort into "immoral" work and then act morally once employed requires further study. Nevertheless, our work suggests an important role for the perceived immorality of work in labor markets.

9. References

Abeler, J., Nosenzo, D., & Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4), 1115-1153.

Akerlof, G. A., and Shiller, R. J. (2015) "Phishing for phools: The economics of manipulation and deception," Princeton University Press.

Andreoni, J. (1989) "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence." Journal of Political Economy, 97(6): 1447–1458.

Andreoni, J., Nikiforakis, N. and Stoop, J. (2021) "Higher socioeconomic status does not predict decreased prosocial behavior in a field experiment." Nature Communications, 12(1): 1-8.

Ariely, D., Bracha, A. and Meier, S. (2009) "Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially," American Economic Review, 99(1): 544-555.

Arunachalam, R. and Shah, M. (2008) "Prostitutes and Brides?," American Economic Review, Papers and Proceedings, 98(2): 516-522.

Ashraf, N., and Bandiera, O. (2017) "Altruistic Capital," American Economic Review, Papers and Proceedings, 107(5): 70-75.

Ashraf, N., Bandiera, O., Davenport, E., and Lee, S. S. (2020) "Losing prosociality in the quest for talent? Sorting, selection, and productivity in the delivery of public services," American Economic Review, 110(5), 1355-1394.

Ashraf, N., Bandiera, O. and Delfino, A. (2020) "The distinctive values of bankers" AEA Papers and Proceedings, 110: 167-171.

Barfort, S., Harmon, N. A., Hjorth, F. and Olsen, A. L. (2019) "Sustaining Honesty in Public Service: The Role of Selection." American Economic Journal: Economic Policy, 11 (4): 96-123.

Bartling, B., Valero, V., and Weber, R. A. (2021) "The causal effect of income growth on consumer social responsibility," working paper.

Bartling, B., Weber, R. A. and Yao, L. (2015) "Do Markets Erode Social Responsibility?," Quarterly Journal of Economics, 130(1): 219-66.

Bénabou, R. and Tirole, J. (2006) "Incentives and Prosocial Behavior," American Economic Review, 96(5): 1652-1678.

Bénabou, R., and Tirole, J. (2011) “Identity, morals, and taboos: Beliefs as assets,” Quarterly Journal of Economics, 126(2): 805-855.

Besley, T. and Ghatak, M. (2005) “Competition and Incentives with Motivated Agents,” American Economic Review, 95(3): 616-636.

Blitz, D. and Fabozzi, F. J. (2017) “Sin Stocks Revisited: Resolving the Sin Stock Anomaly,” Journal of Portfolio Management, 44(1): 1-7.

Bock, O., Baetge, I., and Nicklisch, A. (2014) „Hroot: Hamburg registration and organization online tool,” European Economic Review, 71: 117-120.

British American Tobacco (1998). Corporate social responsibility. Note to Martin Broughton from Heather Honour. Report. URL: https://www.industrydocumentslibrary.ucsf.edu/tobacco/docs/#id=pkgx0195.

British American Tobacco (2015) “Annual Report 2015,” report.

Carpenter, J., and Myers, C. K. (2010) “Why Volunteer? Evidence on the Role of Altruism, Image, and Incentives,” Journal of Public Economics, 94(11–12): 911–20.

Carpenter, J., Matthews, P. and Robbett, A. (2017) “Compensating Differentials in Experimental Labor Markets,” Journal of Behavioral and Experimental Economics, 69: 50-60.

Carter, J. R., and Irons, M. D. (1991) “Are Economists Different, and If So, Why?” Journal of Economic Perspectives, 5(2): 171-177.

Cassar, L. and Meier, S. (2018) “Nonmonetary Incentives and the Implications of Work as a Source of Meaning,” Journal of Economic Perspectives, 32(3): 215-238.

CNBC (2019) “Facebook has struggled to hire talent since the Cambridge Analytica scandal, according to recruiters who worked there,” accessed June 13, 2019, https://www.cnbc.com/2019/05/16/facebook-has-struggled-to-recruit-since-cambridge-analytica-scandal.html.

Cohn, A., Fehr, E. and Maréchal, A. (2014) “Business culture and dishonesty in the banking industry,” Nature, 516: 86-89.

DellaVigna, S., List, J. A., Malmendier, U., and Rao, G. (2016) “Voting to tell others,” Review of Economic Studies, 84(1): 143-181.

Delfgaauw, J. and Dur, R. (2008) “Incentives and Workers’ Motivation in the Public Sector,” Economic Journal, 118(525): 171-191.

Dewatripont, M. and Tirole, J. (2023) “The Morality of Markets,” working paper.

Dufwenberg, M., Heidhues, P., Kirchsteiger, G., Riedel, F., and Sobel, J. (2011) "Other-regarding preferences in general equilibrium," Review of Economic Studies, 78(2), 613-639.

Dur, R. and van Lent, M. (2019) "Socially Useless Jobs," Industrial Relations, 58(3), 543-546.

Dur, R. and Zoutenbier, R. (2014) "Working for a Good Cause," Public Administration Review, 74(2): 144-155.

Edlund L. and Korn, E. (2002) "A Theory of Prostitution," Journal of Political Economy, 110(1): 181-214.

Fehr, E., & Charness, G. (2023). Social preferences: fundamental characteristics and economic consequences. *CESifo Working Paper No. 10488*.

Fehrler, S. and Kosfeld, M. (2014) "Pro-Social Missions and Worker Motivation: An Experimental Study," Journal of Economic Behavior & Organization, 100: 99-110.

Fischbacher, U. (2007) "z-Tree: Zurich toolbox for ready-made economic experiments," Experimental Economics, 10(2), 171-178.

Fisman, R., Jakiela, P., Kariv, S. and Markovits, D. (2015) "The distributional preferences of an elite," Science, 349(6254): 1300.

Frank, R. H. (1996) "What prices the moral high ground?," Southern Economic Journal, 63(1): 1-17.

Frank, R. H., Gilovich, T. and Regan, D. T. (1993) "Does Studying Economics Inhibit Cooperation?," Journal of Economic Perspectives, 7(2): 159-171.

Friebel, G., Kosfeld, M. and Thielmann, G. (2019) "Trust the Police? Self-Selection of Motivated Agents into the German Police Force," American Economic Journal: Microeconomics, 11(4): 59-78.

Gertler, P., Shah, M. and Bertozzi, S. M. (2005) "Risky Business: The Market for Unprotected Commercial Sex," Journal of Political Economy, 113(3): 518-550.

Gill, A., Heinz, M., Schumacher, H., and Sutter, M. (2020) "Trustworthiness in the financial industry," working paper.

Gneezy, U., Rockenbach, B. and Serra-Garcia, M. (2013) "Measuring lying aversion," Journal of Economic Behavior and Organization, 93: 293-300.

Goldstick, J. E., Cunningham, R. M., and Carter, P. M. (2022) "Current causes of death in children and adolescents in the United States," New England journal of medicine, 386(20): 1955-1956.

Hanna, R. and Wang, S. (2017) “Dishonesty and Selection into Public Service: Evidence from India,” American Economic Journal: Economic Policy, 9 (3): 262-290.

Heath, D. (2016) “Contesting The Science of Smoking,” The Atlantic, https://www.theatlantic.com/politics/archive/2016/05/low-tar-cigarettes/481116/

Hedblom, D., Hickman, B. R., and List, J. A. (2019) “Toward an understanding of corporate social responsibility: Theory and field experimental evidence,” working paper.

Hong, H. and Kacperczyk, M. (2009) “The Price of Sin: The Effects of Social Norms on Markets,” Journal of Financial Economics, 93: 15-36.

Hu, J., and Hirsh, J. B. (2017) “Accepting lower salaries for meaningful work,” Frontiers in Psychology, 8: 1649.

Karni, A. (2022) “Assault Weapons Makers Testify They Bear No Responsibility for Gun Violence,” New York Times, https://www.nytimes.com/2022/07/27/us/politics/assault-weapons-revenue.html

Köszegi and Kaufman (2023) “Understanding Markets with Socially Responsible Consumers,” working paper.

Kirchler, M., Huber, J., Stefan, M. and Sutter, M. (2016) “Market design and moral behavior,” Management Science, 62: 2615-2625.

Krueger, P., Metzger, D., and Wu, J. (2023) “The sustainability wage gap,” working paper.

Leete, L. (2001) “Whither the Nonprofit Wage Differential? Estimates from the 1990 Census,” Journal of Labor Economics 19(1): 136-170.

Lockwood, B. B., Nathanson, C. G. and Weyl, E. G. (2017) “Taxation and the Allocation of Talent,” Journal of Political Economy, 125(5): 1635-1682.

Mankiw, G. N. (2010) “Spreading the Wealth Around: Reflections Inspired by Joe the Plumber,” Eastern Economic Journal, 36: 285-298.

Mas, A., and Pallais, A. (2017) “Valuing Alternative Work Arrangements,” American Economic Review, 107(12): 3722-3759.

Merritt, A. C., Effron, D. A., and Monin, B. (2010) “Moral self-licensing: When being good frees us to be bad.” Social and personality psychology compass, 4(5): 344-357.

Mocan, N. H. and Tekin, E. (2003) “Nonprofit Sector and Part-Time Work: An Analysis of Employer-Employee Matched Data on Child Care Workers,” Review of Economics and Statistics 85(1): 38-50.

Murphy, K., Shleifer, A. and Vishny, R. (1991) “The allocation of talent: Implications for growth,” Quarterly Journal of Economics, 106(2): 503-530.

Oh, S. (2021) “Does identity affect labor supply?,” working paper.

Philip Morris International Inc. (1999). PM21 Internal Toolkit. Report. URL: http: //industrydocuments.library.ucsf.edu/tobacco/docs/fgxx0085.

Philip Morris International Inc. (2015) “Form 10-K submitted to US Securities and Exchange Commission,” report.

Rosen, S. (1986) “The theory of equalizing differences,“ Ed: Ashenfelter, O. and Layard, R., in: Handbook of Labor Economics, Elsevier Science Publishers BV.

Sausgruber, R. and Tyran, J-R. (2011) “Are we taxing ourselves? How deliberation and experience shape voting on taxes”, Journal of Public Economics, 95: 164-176.

Smith, A. (1776) “An Inquiry into the Nature and Causes of the Wealth of Nations,” London: W. Strahan and T. Cadell.

Smith, V.L., Williams A.W., Bratton W.K. and Vannoni, M.G. (1982) “Competitive market institutions: Double auctions vs. sealed bid-offer auctions,” American Economic Review, 72(1): 58-77.

Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., and Lerner, J. S. (2000) “The psychology of the unthinkable: taboo trade-offs, forbidden base rates, and heretical counterfactuals.” Journal of personality and social psychology, 78(5): 853.

Tonin, M. and Vlassopoulos, M. (2015) „Corporate Philanthropy and Productivity: Evidence from an Online Real Effort Experiment,“ Management Science, 61(8): 1795-1811.

US Department of Justice, Office of Public Affairs (2016) “Goldman Sachs Agrees to Pay More than $5 Billion in Connection with Its Sale of Residential Mortgage Backed Securities,” https://www.justice.gov/, date accessed: 21.01.2017.

Wiswall, M. and Zafar, B. (2018) “Preference for the Workplace, Investment in Human Capital, and Gender,” Quarterly Journal of Economics, 133(1): 457-507.

Ziegler, A., Romagnoli, G. and Offerman, T. (2020) “Morals in multi-unit markets,” working paper.

Appendix A – Additional Figures and Tables

Figure A1: Timeline of Studies 1 and 2

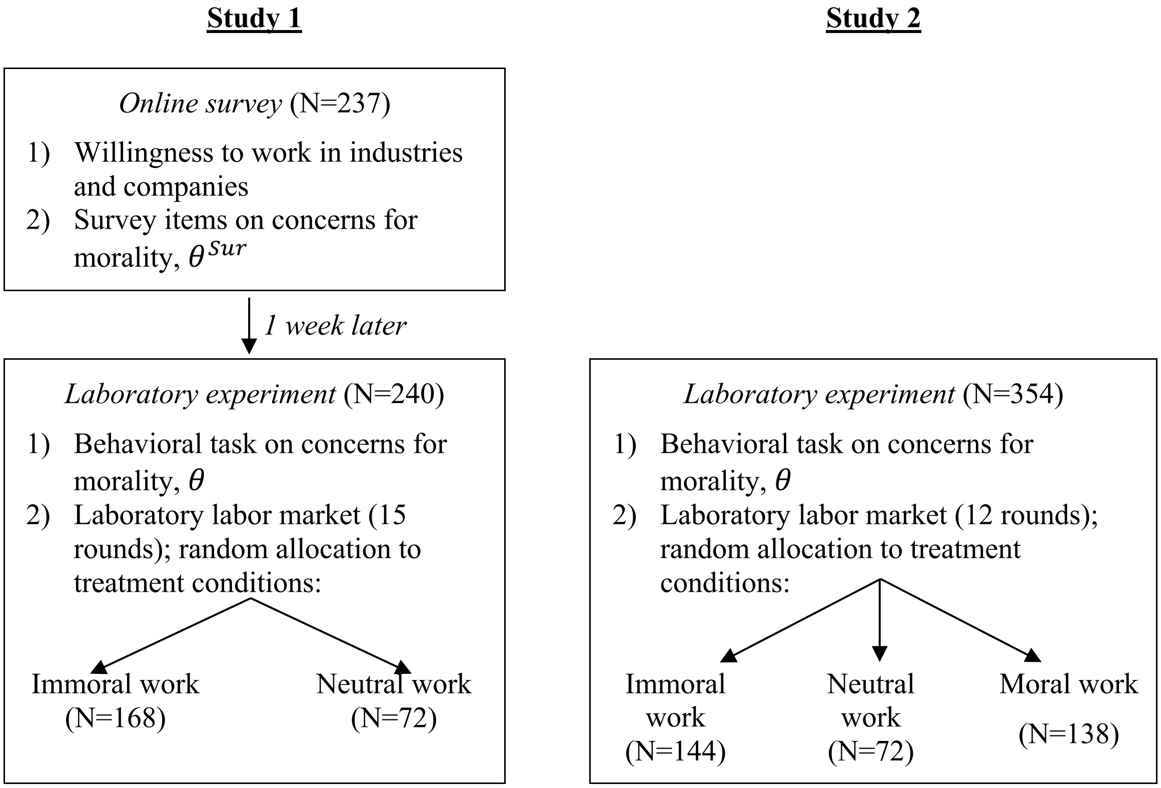
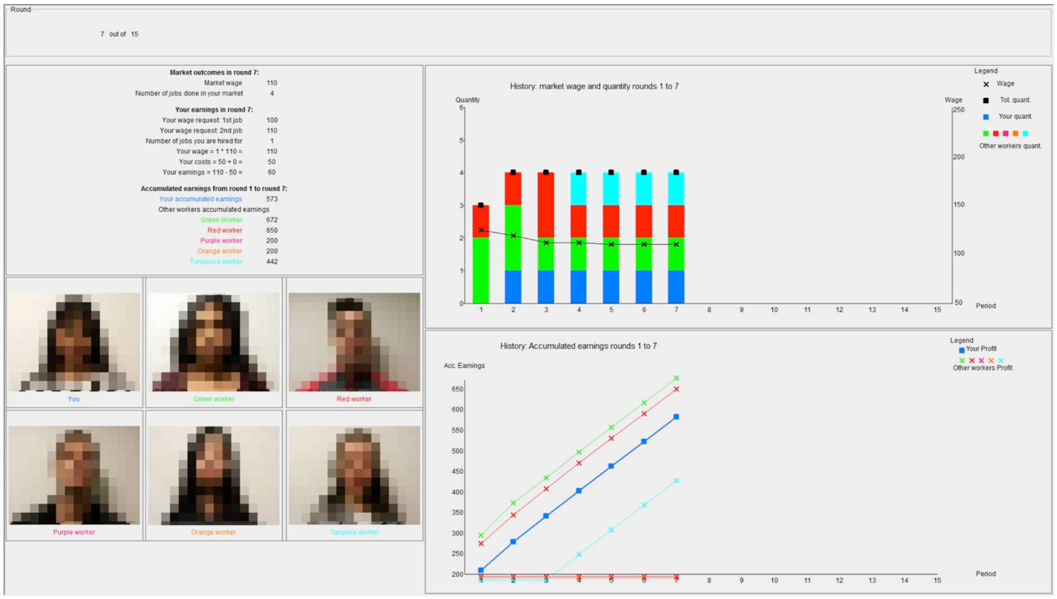

Study 1
Study 2
Online survey (N=237)
1) Willingness to work in industries and companies
2) Survey items on concerns for morality, θ^{Sur}
1 week later
Laboratory experiment (N=240)
1) Behavioral task on concerns for morality, θ
2) Laboratory labor market (15 rounds); random allocation to treatment conditions:
Immoral work (N=168)
Neutral work (N=72)
Laboratory experiment (N=354)
1) Behavioral task on concerns for morality, θ
2) Laboratory labor market (12 rounds); random allocation to treatment conditions:
Immoral work (N=144)
Neutral work (N=72)
Moral work (N=138)

Figure A2: Example of feedback provided after every period (Study 1)



Round

7 out of 15

Market outcomes in round 7:

Market wage	110
Number of jobs done in your market	4

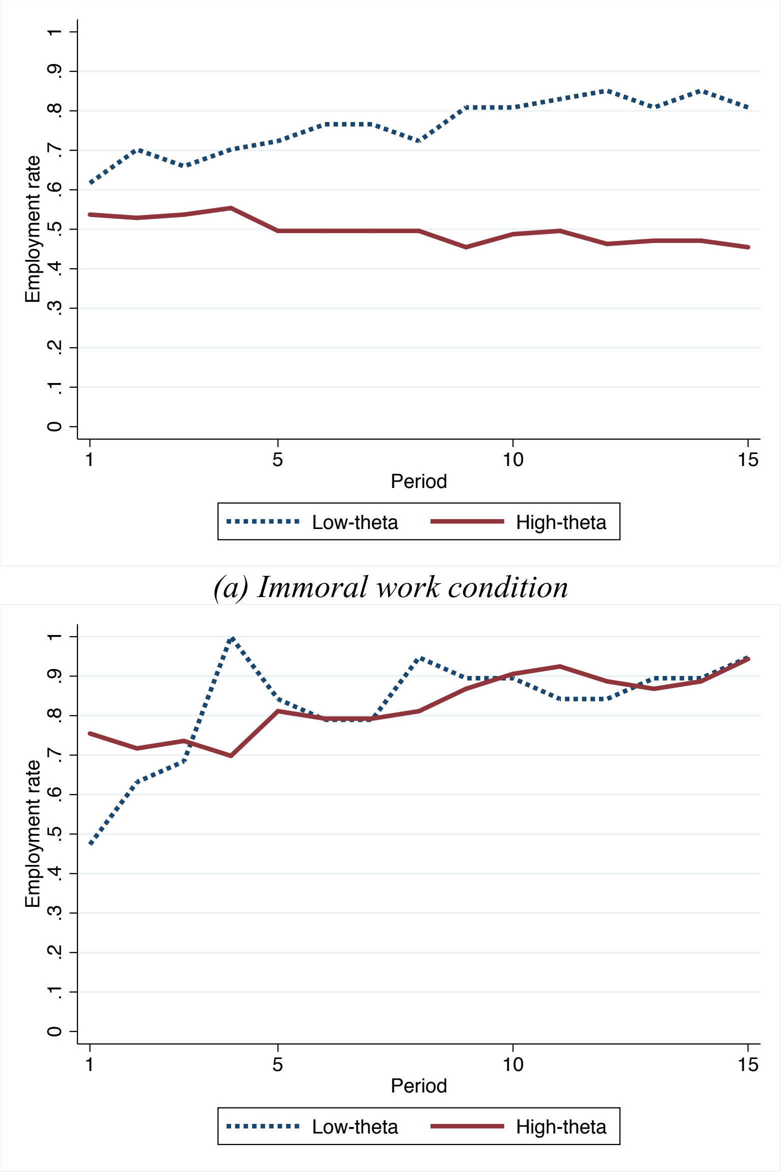
Your earnings in round 7:

Your wage request: 1st job	100
Your wage request: 2nd job	110
Number of jobs you are hired for	1
Your wage = 1 * 110 =	110
Your costs = 50 + 0 =	50
Your earnings = 110 - 50 =	60

Accumulated earnings from round 1 to round 7:

Your accumulated earnings	573
Other workers accumulated earnings	
Green Worker	672
Red worker	650
Purple worker	200
Orange worker	200
Turquoise worker	442

Figure A3: Employment rate by the two moral types (Study 1)



(a) Immoral work condition

(b) Neutral work condition

Figure A4: Labor supply for neutral and immoral work in the laboratory, last 5 periods (Study 1)

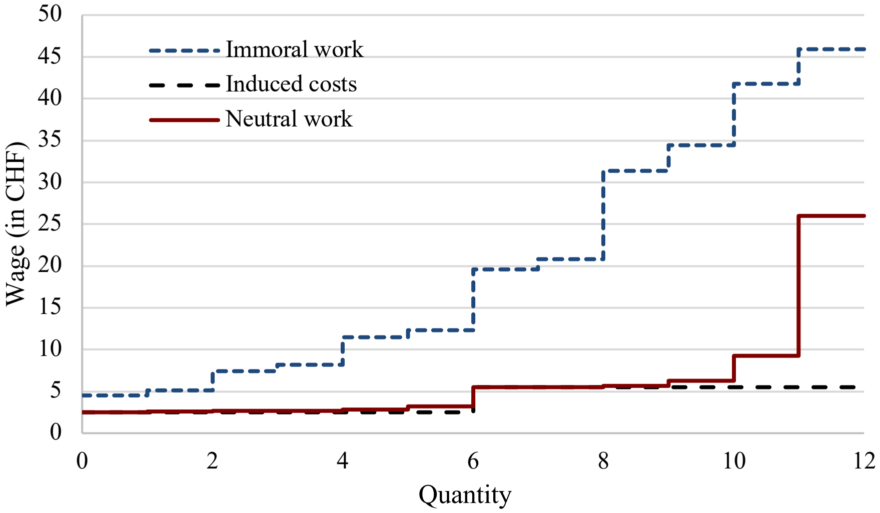
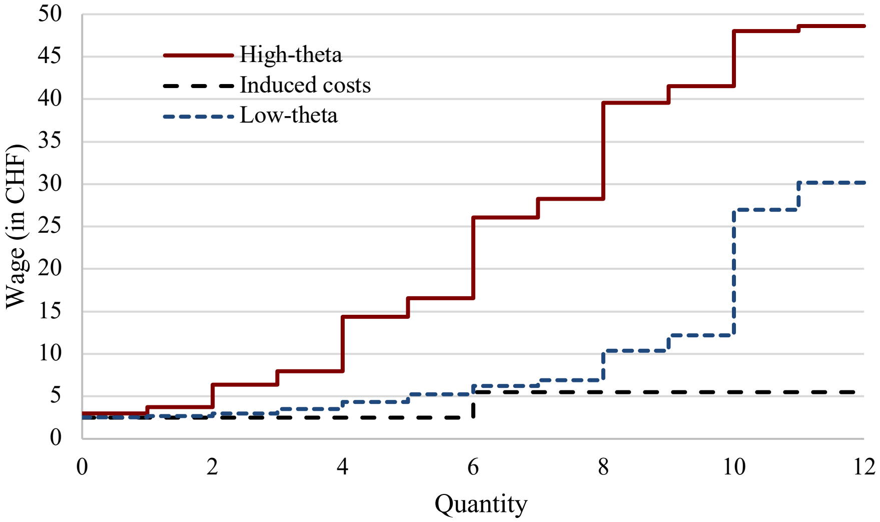


Figure A5: Labor supply for immoral work in the laboratory for different moral types (Study 1)



Notes: Labor supplies conditional on types are calculated with a simulation: 6 labor market decisions (first and second wage request) of high-theta (or, low-theta) types are randomly drawn (without replacement) from our sample. We then calculate the labor supply for this group of people. We repeat this 1000 times (with replacement) and take the average of these 1000 individual labor supplies. This approach differs from the one we use in Figure 3 and Figure A3, where we take the average of the actual labor supplies in the different market groups and periods.

Figure A6: Market quantities in laboratory labor markets (Study 1)

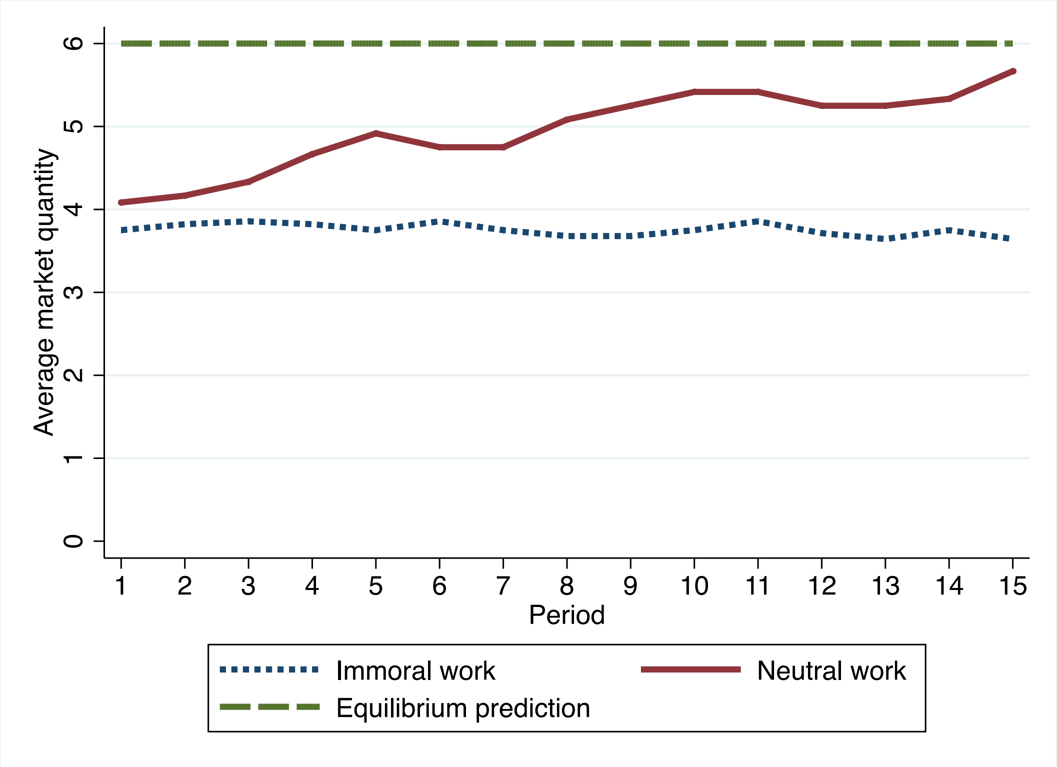
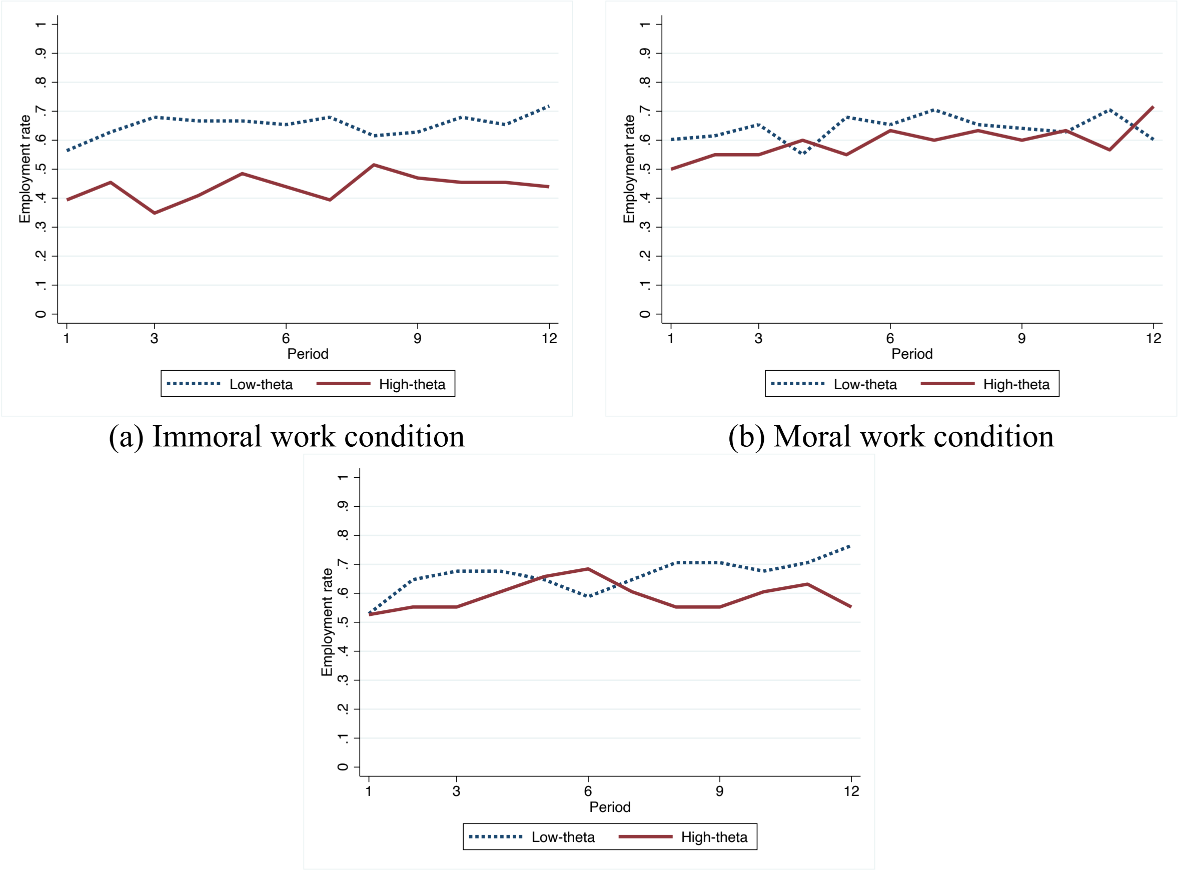


Figure A7: Employment rate by moral type across treatment conditions (Study 2)



(a) Immoral work condition

(b) Moral work condition

(c) Neutral work condition

Figure A8: Costly moral behavior at work (Study 2)

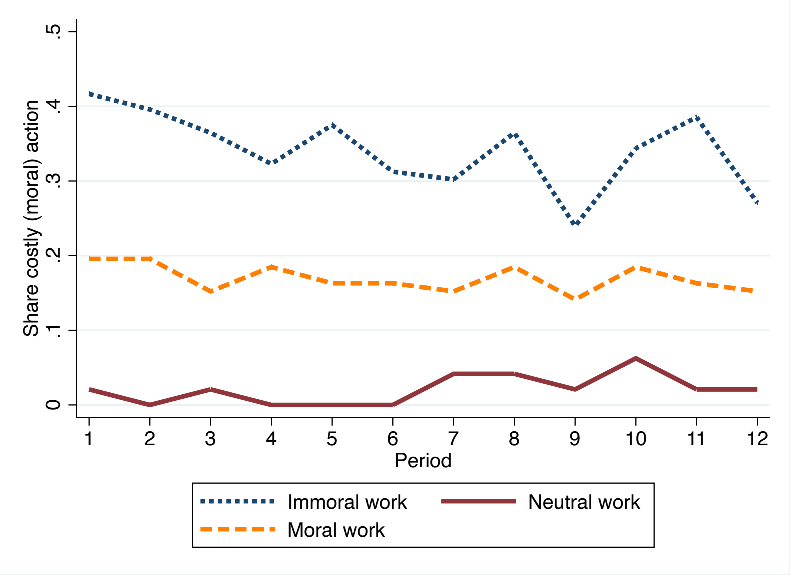
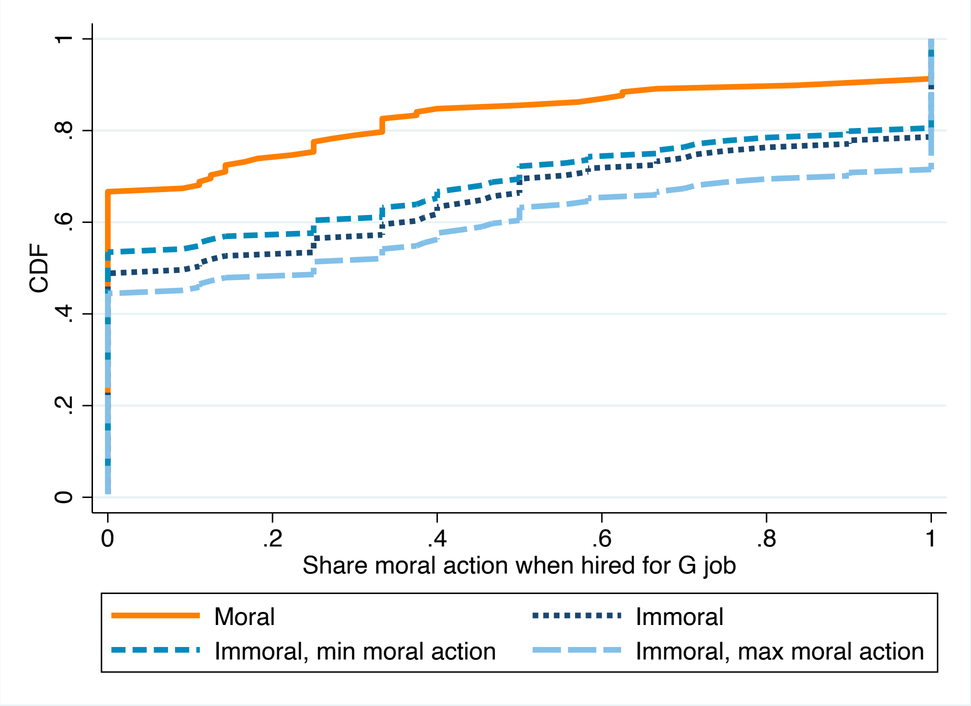


Figure A9: Distribution of costly moral action at work among all workers (Study 2)



Notes: For each participant we calculate the relative frequency with which the participant selected the moral action when hired to do a G job. This figure gives the CDF of these relative frequencies for the immoral and moral work conditions. In the immoral work conditions, 13 participants were never hired for a G job. We do not know the relative frequency for these participants. The line "Immoral" gives the CDF when we exclude these participants. We then provide results for the most extreme cases where we assume that all these participants have a relative frequency of 0 ("Immoral, min moral action") and of 1 ("Immoral, max moral action"). Note that the even when we assume minimal moral behavior, the CDF for immoral work stochastically dominates the CDF for the moral work. We can reject the hypotheses that the relative frequencies are the same in "moral" and "immoral" (t-tests from regressions with standard errors clustered at the market-level: coefficient=-0.174, t=-3.23, p=0.002), in "moral" and "Immoral, max moral action" (coefficient=-0.142, t=-2.25, p=0.029) and in "moral" and "Immoral, min moral action" (coefficient=-0.232, t=-2.90, p=0.006).

Table A1: Relationship between wages and perceived industry immorality

Dependent variable: ln of real gross hourly wage (in 2010 CHF)			
	(1)	(2)	(3)
Perceived industry immorality (standardized)	0.138*** (3.85)	0.072*** (2.77)	0.087*** (4.87)
Age	0.005*** (3.64)	0.007*** (6.94)	0.007*** (6.90)
Male	0.203*** (6.80)	0.155*** (7.62)	0.159*** (6.67)
Married	0.037** (2.13)	0.032 (1.33)	0.031 (1.29)
Education high	0.442*** (8.64)	0.580*** (12.91)	0.572*** (12.95)
Education middle	0.170*** (4.77)	0.288*** (6.60)	0.282*** (6.61)
Swiss	0.036* (1.97)	-0.002 (-0.11)	-0.004 (-0.26)
Experience	0.005*** (3.17)	0.005*** (2.77)	0.005*** (2.84)
Full-time equivalent	-0.038 (-0.54)	-0.168** (-2.44)	-0.168** (-2.28)
Managerial duties	0.065 (0.82)	0.066 (0.93)	0.066 (0.92)
Industry sales	0.034 (1.25)	0.021 (1.22)	0.012 (0.89)
Industry size (employees)	0.001*** (3.69)	0.001*** (4.08)	0.001*** (4.99)
N	32,638	47,935	47,935
Adjusted R^2	0.397	0.263	0.267
Set of industries	First	Second	Second
Sample providing immorality perceptions	Students	Students	Representative
Year and region FE	Yes	Yes	Yes

*Source: Weighed data from the SLFS, years 2010-2016 (wage and demographics), STATENT, years 2011-2016 (industry size, industry sales), Value Added Tax Statistics, years 2010-2016 (industry sales) and our own survey (perceived industry immorality). Notes: Specification (1) uses the immorality perceptions for the initial set of industries (Set of industries = Initial) measured with a student sample (Sample immorality perceptions = Students); specification (2) uses the perceptions for the second set of industries measured with a student sample; specification (3) uses the perceptions for the second set of industries measured with a sample that is representative of the Swiss population. Perceived immorality is standardized. Control variables: Male in {0, 1}, Married in {0, 1}, Education high: higher vocational education and training or university/college, Education middle: apprenticeship, full-time vocational school, matura or pedagogical training, Education low (reference category): compulsory schooling or pre-vocational education, Swiss in {0, 1}, Experience = number of years in the firm, Full-time equivalent = (working hours /42), set to 1 for working hours >= 42, managerial duties in {0, 1}, Industry size = number employees in this industry / 1000 (2010 data is not available, we substitute it with 2011 data), Industry sales = Industry sales/number employees in this industry. All specifications include controls for company region fixed effects (26 Swiss cantons) and year fixed effects (2010-2016). Standard errors clustered at the industry level, t-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*

Table A2: Options available to the "client"

	A	B	C	D	E	F	G	H	I	J
Additional number of children receiving the anti-malarial treatment	1	1	1	-1	1	1	1	1	1	1
Financial reward for client (CHF)	1	1	1	0	1	1	1	1	1	1

Table A3: Distribution of behavior regarding the behavioral measure of concern for morality (Study 1)

Number of lies	Reported number for state r:						Expected payoff lying	Frequency	Share	Classification
	1	2	3	4	5	6				
0 (Honest)	1	2	3	4	5	6	0	161	0.671	High-theta
1	6	2	3	4	5	6	0.83	6	0.038	Low-theta
	2	2	3	4	5	6	0.17	3		
2	6	6	3	4	5	6	1.5	12	0.104	Low-theta
	1	2	3	6	6	6	0.5	2		
	1	3	3	5	5	6	0.33	1		
	1	4	4	4	5	6	0.5	1		
	5	6	3	4	5	6	1.33	2		
	6	5	3	4	5	6	1.33	1		
	3	2	3	5	5	6	0.5	1		
	3	3	3	4	5	6	0.5	5		
3	6	6	6	4	5	6	2	11	0.050	Low-theta
	4	2	3	6	6	6	1	1		
4	6	6	6	6	5	6	2.33	3	0.017	Low-theta
	6	5	5	5	5	6	1.83	1		
5	6	6	6	6	6	6	2.5	15	0.067	Low-theta
	2	3	4	5	6	6	0.83	1		
Lied in a self-harmful manner	1	2	3	4	3	3	-0.83	1	0.054	High-theta
	1	2	3	4	4	4	-0.5	1		
	1	2	3	4	5	5	-0.17	1		
	1	3	2	5	4	6	0	1		
	1	4	2	4	5	6	0.17	1		
	1	4	6	3	5	6	0.67	1		
	2	1	3	4	5	6	0	1		
	3	4	5	4	6	2	0.5	1		
	5	1	3	6	4	2	0	1		
	5	2	3	4	1	6	0	1		
	5	4	6	4	6	5	1.5	1		
	6	2	5	5	1	3	0.17	1		
	6	6	6	6	6	5	2.33	1		

Notes: Expected payoff from lying = $\frac{1}{6}\sum_{r=1}^{6}(m_{ir}-r)$*, where* m_{ir} *is the number that individual* i *reports if the actual die roll is* r.

Table A4: Relationship between participation decision/reservation wage and θ (Hurdle model, Study 1)

Dependent variable:	Participate	Reservation wage	Reservation wage
	(1)	(2)	(3)
Low-theta ($\boldsymbol{\theta_L}$)	0.925***	-0.362	-0.060
	(4.64)	(-0.99)	(-0.39)
Period (t)	-0.019**	-0.047	-0.050***
	(-2.40)	(-1.64)	(-5.74)
Period * $\boldsymbol{\theta_L}$	0.012	-0.015	0.005
	(1.14)	(-0.40)	(0.54)
Constant	0.449***	4.425***	3.312***
	(3.86)	(14.76)	(25.63)
Sigma		2.630***	0.571***
		(7.69)	(10.58)
Condition	Immoral work	Immoral work	Neutral work
N	2,520	1,755	1,077
p-value: t + t* $\boldsymbol{\theta_L^{Exp}}$= 0	0.422	0.001	0.0000

*Notes: Estimates from Craggs double-hurdle model: (1) is a probit model; (2) and (3) are truncated linear regressions (truncated from above at 50 CHF). Models (1) and (2) use only data from the immoral work condition; model (3) uses only data from the neutral work condition. For neutral work, we do not report the regression of market participation as we have only 3 incidences where a subject did not participate. Independent variables: Low-theta in {0, 1}, Period between 1 and 15. Standard errors clustered at market level; z-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

Table A5: Relationship between the behavioral measures of concern for morality and outcomes in the experimental labor markets, robustness (Study 1)

Dependent variable:	**Employment rate**			
	(1)	(2)	(3)	(4)
Number of lies				
1 lie	0.201		0.004	
	(1.07)		(0.11)	
2 lies	0.220**		-0.011	
	(2.21)		(-0.22)	
3 lies	0.398***		-0.296***	
	(5.38)		(-13.43)	
4 lies	0.392***		0.171***	
	(5.03)		(7.74)	
5 lies	0.286***		0.037	
	(3.32)		(0.51)	
self-harmful lies	0.252**		-0.0516	
	(2.57)		(-1.31)	
Expected payoff lying		0.127***		0.006
		(5.05)		(0.18)
Constant	0.475***	0.510***	0.829***	0.824***
	(11.80)	(13.83)	(37.63)	(35.89)
Condition	Immoral	Immoral	Neutral	Neutral
N	2,520	2,520	1,080	1,080
R^2	0.0790	0.0491	0.013	0.0001

Notes: Coefficient estimates of linear regression models. Models (1) and (2) use only data from the immoral work condition, models (3) and (4) use only data from the neutral work condition. Expected payoff from lying = $\frac{1}{6}\sum_{r=1}^{6} m_{ir} - r$ *(*$\in [-2.5,2.5]$*), where* m_{ir} *is the number that individual* i *reports if the actual die roll is* r*. Standard errors clustered at market level; t-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

Table A6: Relationship between θ and behaviors and outcomes in the experimental labor markets (Study 1)

Dependent variable:	**Participation**	**Reservation wage**	**Employment rate**		**Number of work units**	
	(1)	(2)	(3)	(4)	(5)	(6)
Low-theta (θ_L)	1.027***	-0.509*	0.263***	0.214***	0.252***	0.231***
	(4.73)	(-1.81)	(5.54)	(3.76)	(4.95)	(3.74)
Neutral work (N)		-1.114***	0.326***		0.267***	
		(-5.28)	(7.67)		(7.08)	
θ_L * N		0.385	-0.241***	-0.204***	-0.229***	-0.221**
		(1.19)	(-3.74)	(-2.75)	(-3.20)	(-2.57)
Constant	0.329	3.016***	0.627***		0.665***	
	(0.463)	(4.28)	(5.19)		(5.58)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Market FE	No	No	No	Yes	No	Yes
N	2,475	2,788	3,555	3,555	3,555	3,555
LL (pseudo)	-1405.8	-6035.4	-	-	-	-
R^2	-	-	0.116	0.200	0.069	0.112
p-value: θ_L + θ_L*N = 0	-	0.322	0.601	0.839	0.591	0.837

*Notes: Models (1) and (2): Estimates from Craggs double-hurdle Model: (1) probit model; (2) truncated linear regressions (truncated from above at 50 CHF). Note model (1) uses only data from the immoral work condition because non-participation is virtually non-existent for neutral work. Moreover, we do not control for market fixed effects because this introduces technical problems in the double-hurdle model for markets in which all participants participated. Models (3) to (6): Estimates from linear regression models. Independent variables: Low-theta in {0, 1}, Neutral work in {0, 1}. Note that column "p-value: θ_L + θ_L*N = 0" test sorting into neutral work. Control variables: age, gender, Swiss nationality. Standard errors clustered at market level; t-statistics in parentheses; * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.*

Table A7: Distribution of behavior regarding the behavioral measure of concern for morality (Study 2)

Number of lies	Reported number for state r:						Expected payoff lying	Frequency	Share	Classification
	1	2	3	4	5	6				
0 (Honest)	1	2	3	4	5	6	0	136	0.384	High-theta
1	6	2	3	4	5	6	0.83	12	0.040	Low-theta
	5	2	3	4	5	6	0.67	1		
	1	2	3	4	6	6	0.17	1		
2	6	6	3	4	5	6	1.5	22	0.110	Low-theta
	1	2	3	6	6	6	0.5	3		
	1	2	4	4	6	6	0.33	2		
	1	4	3	5	5	6	0.5	1		
	2	2	3	4	6	6	0.33	1		
	3	3	3	4	5	6	0.5	3		
	4	5	3	4	5	6	1	1		
	5	4	3	4	5	6	1	1		
	5	6	3	4	5	6	1.33	2		
	6	5	3	4	5	6	1.33	3		
3	6	6	6	4	5	6	2	43	0.167	Low-theta
	1	2	6	6	6	6	1	1		
	4	4	4	4	5	6	1	1		
	4	5	3	4	6	6	1.17	1		
	4	5	6	4	5	6	1.5	4		
	4	6	4	4	5	6	1.33	1		
	4	6	5	4	5	6	1.5	1		
	5	5	5	4	5	6	1.5	3		
	6	5	4	4	5	6	1.5	1		
	6	5	6	4	5	6	1.83	1		
	6	6	3	6	5	6	1.83	1		
	6	6	4	4	5	6	1.67	1		
4	6	6	6	6	5	6	2.33	11	0.045	Low-theta
	1	6	6	6	6	6	1.67	2		
	4	5	6	6	5	6	1.83	1		
	5	6	5	6	5	6	2	1		
	6	6	5	5	5	6	2	1		
5	6	6	6	6	6	6	2.5	60	0.175	Low-theta
	5	5	5	6	6	6	2	2		

See next page for the rest of the table.

Number of lies	Reported number for state r:						Expected payoff lying	Frequency	Share	Classification
	1	2	3	4	5	6				
Lied in a self-harmful manner	1	1	1	1	1	1	-2.5	1	0.079	High-theta
	1	2	4	3	5	6	0	1		
	1	3	4	3	5	4	-0.17	1		
	1	4	3	6	5	4	0.33	1		
	1	4	5	4	1	6	0	1		
	1	4	5	6	3	2	0	1		
	1	4	6	2	5	3	0	1		
	1	5	4	2	6	3	0	1		
	2	1	3	4	6	5	0	1		
	2	4	5	4	2	6	0.33	1		
	3	2	1	6	5	4	0	1		
	3	2	4	5	1	6	0	1		
	3	4	3	4	3	4	0	1		
	3	5	1	4	2	6	0	1		
	3	6	1	5	4	2	0	1		
	4	1	2	3	4	5	-0.33	1		
	4	4	4	4	4	4	0.5	1		
	4	5	3	6	5	4	1	1		
	5	2	6	4	5	3	0.67	1		
	5	2	6	6	3	4	0.83	1		
	5	3	6	2	1	4	0	1		
	5	4	3	4	6	2	0.5	1		
	5	4	6	3	2	1	0	1		
	5	5	5	5	5	5	1.5	1		
	5	6	5	4	5	3	1.17	1		
	6	4	4	3	3	2	0.17	1		
	6	5	6	4	1	2	0.5	1		
	6	6	6	6	4	5	2	1		

Notes: Expected payoff from lying = $\frac{1}{6}\sum_{r=1}^{6}(m_{ir}-r)$*, where* m_{ir} *is the number that individual* i *reports if the actual die roll is* r*.*

Table A8: Relationship between the behavioral measures of concern for morality and outcomes in the experimental labor markets, robustness (Study 2)

Dependent variable:	**Employment rate**					
	(1)	(2)	(3)	(4)	(5)	(6)
Number of lies						
1 lie	0.079		-0.044		-0.174*	
	(0.55)		(-0.49)		(-1.94)	
2 lies	0.107		0.040		-0.028	
	(1.53)		(0.83)		(-0.37)	
3 lies	0.179**		0.068		0.028	
	(2.54)		(1.13)		(0.33)	
4 lies	0.312***		0.055		0.284***	
	(3.61)		(0.71)		(10.59)	
5 lies	0.312***		0.100*		0.053	
	(4.83)		(2.00)		(0.98)	
self-harmful	0.002		0.107*		-0.178	
lies	(0.02)		(1.78)		(-1.42)	
Expected payoff		0.112***		0.026		0.036
lying		(5.24)		(1.39)		(1.71)
Constant	0.438***	0.435***	0.580***	0.594***	0.632***	0.592***
	(12.76)	(14.75)	(20.62)	(28.85)	(23.54)	(31.35)
Condition	Immoral	Immoral	Moral	Moral	Neutral	Neutral
N	1,728	1,728	1,656	1,656	864	864
R^2	0.064	0.059	0.009	0.003	0.037	0.006

Notes: Coefficient estimates of linear regression models. Models (1) and (2) use only data from the immoral work condition, models (3) and (4) use only data from the moral work condition, models (5) and (6) use only data from the neutral work condition. Expected payoff from lying = $\frac{1}{6}\sum_{r=1}^{6} m_{ir} - r$ *(*$\in [-2.5,2.5]$*), where* m_{ir} *is the number that individual* i *reports if the actual die roll is* r*. Standard errors clustered at market level; t-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

Table A9: Relationship between θ and behaviors and outcomes in the experimental labor markets, robustness (Study 2)

Dependent variable:	**Participation**	**Reservation wage**	**Employment rate**		**Number of work units**	
	(1)	(2)	(3)	(4)	(5)	(6)
Low-theta ($\boldsymbol{\theta_L}$)	0.985***	-0.826***	0.209***	0.223***	0.186**	0.255**
	(4.02)	(-2.76)	(4.06)	(3.36)	(2.41)	(2.62)
Moral work (M)	0.967***	-1.143***	0.161***	-	0.088*	-
	(4.39)	(-4.11)	(3.96)		(1.82)	
Neutral work (N)	1.190***	-0.651**	0.150***	-	0.075	-
	(5.10)	(-2.02)	(3.75)		(1.57)	
$\boldsymbol{\theta_L}$ * M	-0.828**	0.743**	-0.183***	-0.186**	-0.169*	-0.234**
	(-2.47)	(2.38)	(-2.85)	(-2.29)	(-1.94)	(-2.16)
$\boldsymbol{\theta_L}$ * N	-1.159***	0.648	-0.154**	-0.166*	-0.145	-0.209*
	(-2.81)	(1.40)	(-2.19)	(-1.91)	(-1.57)	(-1.83)
Constant	1.110*	2.503***	0.611***	-	0.651***	-
	(1.68)	(5.40)	(5.37)		(4.60)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Market FE	No	No	No	Yes	No	Yes
N	4,248	3,985	4,248	4,248	4,248	4,248
LL (pseudo)	-836.4	-8920.0	-	-	-	-
R^2	-	-	0.033	0.044	0.013	0.017
p-value: $\boldsymbol{\theta_L}$ + $\boldsymbol{\theta_L}$*M = 0	0.534	0.399	0.502	0.414	0.667	0.665
p-value: $\boldsymbol{\theta_L}$ + $\boldsymbol{\theta_L}$*N = 0	0.609	0.616	0.285	0.338	0.471	0.500
p-value: M = N	0.351	0.017	0.739	-	0.713	-
p-value: $\boldsymbol{\theta_L}$*M = $\boldsymbol{\theta_L^{Exp}}$*N	0.425	0.798	0.640	0.782	0.728	0.764

*Notes: Models (1) and (2): Estimates from Craggs double-hurdle Model: (1) probit model; (2) truncated linear regressions (truncated from above at 50 CHF). Note that we do not control for market fixed effects because this introduces technical problems in the double-hurdle model for markets in which all participants participated. Models (3) to (6): Estimates from linear regression models. Specification (3) is the pre-registered main specification. Independent variables: Low-theta in {0, 1}, Moral work in {0, 1}, Neutral work in {0, 1}. Note that columns "p-value: $\theta_L^{\square}$ + $\theta_L^{\square}$*M = 0" and "p-value: $\theta_L^{\square}$ + $\theta_L^{\square}$*N = 0" test sorting into moral and neutral work, respectively. Columns "p-value: M = N" and "p-value: $\theta_L^{\square}$*M = $\theta_L^{\square}$*N" test treatment differences in sorting between the moral and neutral work conditions. Control variables: age, gender, Swiss nationality, Chinese nationality. Standard errors clustered at market level; t-statistics in parentheses; * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.*

Table A10: Perceived immorality of firms

Firms	Perceived immorality I(j)	Firms	Perceived immorality I(j)
Marlboro	0.54	Swisscom	-0.07
Monsanto	0.52	Firmenich	-0.09
Glencore	0.46	Winterthur Assurance	-0.1
Philip Morris	0.46	Swiss Life	-0.13
Nestlé	0.39	Swatch	-0.17
Tamoil	0.37	Adecco	-0.18
Syngenta	0.23	ABB	-0.2
UBS	0.19	Migros	-0.38
Novartis	0.18	WWF	-0.66
Credit Suisse	0.17	Pro Juventute	-0.66
Roche	0.13	Pro Natura	-0.67
Holcim	0.03	UNICEF	-0.72
Ernst and Young	-0.05	Red cross	-0.81

Table A11: Regressions of willingness to work for diverse industries and firms on perceived immorality and moral types (Study 1)

Dependent variable:	**Willingness to work for industry *j***		**Willingness to work for firm *j***	
	(1)	(2)	(3)	(4)
Perceived immorality (I(j))	-0.232*** (-3.99)	-0.226*** (-4.72)	-0.140** (-2.08)	-0.139** (-2.07)
Moral Type ($\boldsymbol{\theta_H}$)	-0.027 (-1.34)	-0.029 (-1.49)	-0.043 (-1.58)	-0.051** (-1.96)
$\boldsymbol{\theta_H}$ * I(j)	-0.078*** (-2.56)	-0.078** (-2.51)	-0.154*** (-4.30)	-0.154*** (-4.38)
N	4715	4715	5064	5064
Control variables	No	Yes	No	Yes

*Notes: Coefficient estimates of linear regression models. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: We use θ_H to classify participants, where $\theta_H^{\square}$=0 for low-theta types and $\theta_H^{\square}$=1 for high-theta types. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study, average wage industry 2016 (SLFS; only for industries), industry size 2016 (STATENT; only for industries), industry sales 2015 (Value Added Tax Statistics; only for industries). Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*

Table A12: Regressions of willingness to work for diverse industries and firms on perceived immorality and moral types, robustness checks classification firms' immorality (Study 1)

Dependent variable:	**Willingness to work for firm *j***	
	(1)	(2)
Perceived immorality ($\mathbf{I}_{\mathrm{Alt}}$(j))	-0.157** (-1.96)	-0.156** (-1.96)
Moral type ($\boldsymbol{\theta_H}$)	-0.045* (-1.65)	-0.053** (-2.04)
$\boldsymbol{\theta_H}$ * $\mathbf{I}_{\mathrm{Alt}}$(j)	-0.174*** (-4.06)	-0.175*** (-4.15)
N	5'064	5'064
Control variables	No	Yes

*Notes: Coefficient estimates of linear regression models. Perceived immorality is calculated different then in our main analysis: Clients that choose "I don't know this organization" are classified as giving neutral ratings. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: we use θ^{Exp} to classify participants, where θ_H=0 for low-theta types and θ_H=1 for high-theta types. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study. Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*

Table A13: Regressions of willingness to work for diverse industries and firms on perceived immorality and moral types, robustness checks for "I don't know this organization" (Study 1)

Dependent variable:	**Willingness to work for firm *j***			
	(1)	(2)	(3)	(4)
Perceived immorality (I(j))	-0.123**	-0.123**	-0.092	-0.092
	(-2.21)	(-2.20)	(-1.49)	(-1.46)
Moral type ($\boldsymbol{\theta_H}$)	-0.034	-0.040*	-0.039	-0.056
	(-1.54)	(-1.91)	(-0.58)	(-0.90)
$\boldsymbol{\theta_H}$ * I(j)	-0.131***	-0.131***	-0.158***	-0.158***
	(-4.47)	(-4.43)	(-2.98)	(-2.87)
N	6'162	6'162	1'352	1'352
Control variables	No	Yes	No	Yes

*Notes: Coefficient estimates of linear regression models. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not fill out the questionnaire are excluded. Columns (1), (2): Observations where subjects did not know the firm ("I don't know this organization") are classified as having willingness to work of 0.5. Columns (3), (4): only participants that did know all firms (N=52) are included. Independent variables: We use θ_H to classify participants, where θ_H=0 for low- theta types and θ_H=1 for high-theta types. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study. Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

Table A14: Regressions of willingness to work for diverse industries and firms on employment rate in the immoral work condition (Study 1)

Dependent variable:	**Willingness to work for industry *j***		**Willingness to work for firm *j***	
	(1)	(2)	(5)	(6)
Perceived immorality (I(j))	-0.332***	-0.399***	-0.311***	-0.310***
	(-6.64)	(-6.74)	(-6.42)	(-6.33)
Employment rate (E)	0.029	0.044*	0.048	0.055
	(1.07)	(1.79)	(1.36)	(1.54)
E * I(j)	0.073**	0.073**	0.114**	0.114***
	(2.03)	(2.00)	(2.49)	(2.44)
N	3'275	3'275	3'561	3'561
Control variables	No	Yes	No	Yes

*Notes: Coefficient estimates of linear regression models. Sample incudes only subjects from the immoral work condition. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: Employment rate is the share of market periods in which the worker was employed. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study, average wage industry 2016 (SLFS; only for industries), industry size 2016 (STATENT; only for industries), industry sales 2015 (Value Added Tax Statistics; only for industries). Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

Appendix B – Details of analysis using Swiss Labor Force Survey

In this section, we provide additional details of the analysis of the Swiss labor market data discussed in Section 4. We investigate whether the perception that an industry involves immoral work is associated with a wage premium. To do so, we obtain novel measures of the perceptions of the immorality of various industries in Switzerland, which we compare with workers' wages from the Swiss Labor Force Survey (SLSF). Our general approach for obtaining ratings of perceived industry immorality proceeds in two steps: (i) selection of a broad set of industries and (ii) independent ratings of perceived industry (im)morality. We employ different methods for these two steps.

In Section B.1 we discuss our first approach. In step (i), we identified industries that we jointly perceived as involving work activities likely to be widely seen as immoral. In step (2), we then obtained independent ratings of the perceived immorality of these industries through a survey of university students in Switzerland.

The approach discussed in Sections B.2 and B.3, implemented step (i) by asking research assistants who were not familiar with the research question to identify industries. We then collect immorality ratings for this second set of industries with two samples: a new sample of students (Section B.2) and a larger survey sample broadly representative of the German- and French-speaking population of Switzerland (Section B.3).

Finally, in Section B.4, we do not use a measure of the perceived industry immorality, but instead focus on wages in a set of "sin industries."

B.1 Initial set of industries, student ratings

Set of industries: We initially identified industries that we jointly perceived as involving work activities likely to be widely seen as immoral; we did so before looking at any data from these industries, including wages.[1] This yielded six "immoral" industries: gambling and betting activities, monetary intermediations, credit granting, manufacture of tobacco products, wholesale of tobacco products and manufacture of weapons and ammunition. These include the industries regularly classified as "sin industries" in financial research (e.g., Hong and Kacperczyk, 2009;

[1] Specifically, we started with the complete list of industries listed in the SLFS. Each of the authors went through the list and indicated any industries that he or she believed was widely perceived to have a significant immoral component. We selected those industries for which all three authors agreed. We proceed this way, rather than using the entire set of Swiss industries, to keep the number of questions we ask in subsequent surveys manageable.

Blitz and Fabozzi, 2017). We then chose comparison industries from within the same industrial branch with similar distributions of education levels, as well as nine additional industries representing large shares of employment in Switzerland.[2] Table B1 gives all selected industries.

Industry ratings: We next obtained independent ratings of the perceived (im)morality of the selected industries. We asked a sample of 177 students on the campus of the University of Zurich and the Federal Institute of Technology (ETH) to rate each industry on a 5-point Likert scale ranging from "very moral" to "very immoral," re-scaled the responses to lie on the -1 to 1 interval and averaged the responses. (These survey data were collected as part of our survey studies, which we describe in more detail in Section 5 of the paper.) Table B1 gives the ratings for all industries.

Results: Figure 1 (i) in the main text shows the correlation between average industry wages and the measure of perceived industry immorality. Table B2 reports regressions of the natural logarithm of real gross hourly wages on the new collected measure of perceived industry immorality, along with several additional control variables.[3] We find a substantial and statistically significant immorality wage premium. According to Model 3, a one standard deviation increase in perceived industry immorality is associated with a 14.8 percent higher (geometric) mean hourly wage (z = 3.59, p < 0.001, 95%-CI: [6.7, 22.9]).[4]

B.2 Second set of industries, student ratings

While we did not look at wages when selecting the industries, a natural concern is that choosing the sample of industries ourselves possibly (unconsciously) biases the sample toward those likely to confirm our hypothesis. As a robustness check, we elicit the perceived industry immorality for a second set of industries, but in this case we do not select the industries. Instead, we ask research assistants unaware of our hypotheses to select moral and immoral industries. We then provide evidence for an immorality premium in this second set of industries.

[2] We chose five comparison industries: non-life insurance (for monetary intermediations; credit granting), organization and operation of sport facilities (for gambling and betting activities), processing of tea and coffee (for manufacture of tobacco products), manufacture of electronic components (for manufacture of weapons and ammunitions), wholesale of perfume and cosmetics (for wholesale of tobacco products). These twenty industries make up a substantial share of the Swiss labor market: they employ 20.6% of the Swiss labor force (STATENT, 2016).

[3] Note that numbers of observations differ substantially between industries. If we weight observations by industry size (instead of using survey weights), estimates for perceived immorality are smaller (Model 1: 0.106, Model 2: 0.076, Model 3: 0.074), but still significant (t=3.88, 3.10, and 4.03, respectively).

[4] We obtain this number by doing the following calculation: $e^{0.138} - 1 \approx 0.148$. We use the delta method to calculate the corresponding z-values, p-values and confidence intervals (CIs).

Set of industries: We created a list with all industries that had at least 50 observations in the Swiss Labor Force Survey. This resulted in a list of 394 industries. We then asked five research assistants to select up to ten industries in which they think it is most immoral to work and up to ten industries in which they think it is most moral to work. They ranked the selected industries from most immoral (moral) to least immoral (moral). The research assistants were unaware of our research question. The five industries that the research assistants thought to be the most immoral and the five industries that they thought to be the most moral are included in the set of industries used in this robustness test.[5] In addition to these ten industries, we randomly selected a set of 40 other industries that had at least 50 observations in the Swiss Labor Force Survey. Table B3 gives all selected industries.

Industry ratings: We elicited a measure of perceived immorality for the set of 50 industries. We recruited 45 participants drawn from the same subject pool from which we recruit participants for our laboratory experiment (but that did not participate in our experiment). These participants rated how immoral they think it is to work for each of the 50 industries on a 7-point Likert scale ranging from very immoral to very moral. We re-scaled the responses to lie on the -1 to 1 interval. In Table B3, column "Perc. Imm. B2" gives the ratings for all industries.

Results: Figure 1 (ii) shows the correlation between average industry wages and the new collected measure of perceived industry immorality. Table B4 reports regressions of the natural logarithm of real gross hourly wages on the new collected measure of perceived industry immorality, along with several additional control variables (Model 1 to 3).[6] We find a substantial and statistically significant immorality wage premium. According to Model 3, we estimate that a one standard deviation increase in perceived industry immorality is associated with a 7.4 percent percent higher (geometric) mean hourly wage (z = 2.67, p = 0.008, 95%-CI: [2.0, 12.9]).[7]

[5] We calculated the average rank of each industry as follows. We first allocated points to each industry according to its rank: if an in industry was rated the most immoral industry it received 10 points, the second most immoral industry received 9 points, and so on. We then added up points for every industry and selected the five immoral and the five moral industries with the highest number of points.

[6] Numbers of observations differ substantially between industries. If we weight observations by industry size (instead of using survey weights), estimates for perceived immorality are similar (Model 1: 0.117, Model 2: 0.096, Model 3: 0.091) and statistically significant (t=2.83, t=3.21,t=3.39, respectively).

[7] We obtain this number by doing the following calculation: $e^{0.072} - 1 \approx 0.074$.

B.3 Second set of industries, ratings from representative sample

In the analysis in section B.1 and B.2, we rely on student populations to measure perceived industry immorality. As a third robustness check, we elicit the perceived industry immorality from a representative sample of the Swiss population.

Set of industries: We use the set of 50 industries from section B.2 that were selected by research assistants who were not familiar with the research question.

Industry ratings: We elicited a measure of perceived immorality in April 2023. We recruited a sample of 303 participants that is broadly representative (in terms of age, gender and region) of the Swiss population in the German- and French-speaking regions of Switzerland with help of the survey company LINK. LINK actively recruits people to create a representative panel in terms of age, gender, and region. Table B5 gives the summary statistics of the sample. Instructions were available in both German and French. These participants rated how immoral they think it is to work for each of the industries on a 7-point Likert scale ranging from very immoral to very moral. We re-scaled the responses to lie on the -1 to 1 interval. Table B3 column "Perc. Imm. B3" gives the ratings for all industries. While students tend to view the meat processing industry as more immoral than the general population, the overall immorality ratings are similar. The correlation between the immorality ratings given by students and the general Swiss population is high (corr = 0.91).

Results: Figure 1 (iii) shows the correlation between average industry wages and the new collected measure of perceived industry immorality. Table B4 reports regressions of the natural logarithm of real gross hourly wages on the new collected measure of perceived industry immorality, along with several additional control variables (Models 4 and 6).[8] We find a substantial and statistically significant immorality wage premium. According to Model 6, we estimate that a one standard deviation increase in perceived industry immorality is associated with a 9.1 percent higher (geometric) mean hourly wage ($z = 4.66$, $p < 0.001$, 95%-CI: [5.3, 13.0]).[9]

[8] Numbers of observations differ substantially between industries. If we weight observations by industry size (instead of using survey weights), estimates for perceived immorality are similar (Model 4: 0.125, Model 5: 0.100, Model 6: 0.092) and statistically significant (t=3.89, t=3.97,t=3.90, respectively).

[9] We obtain this number by doing the following calculation: $e^{0.087} - 1 \approx 0.091$.

B.4 Industry dummies instead of ratings

In a final robustness check, we use all non-immoral industries in the Swiss Labor Force survey as control industries. We do not have a measure of the perceived industry immorality, $I(j)$, for most industries. Instead of relying on such a measure, we define a set of industries as "immoral industries" and calculate wage premiums (or, wage discounts) for these industries, controlling for worker, job and industry characteristics. This approach is commonly used in the literature that studies stock returns for "sin industries" (e.g., Hong and Kacperczyk, 2009; Blitz and Fabozzi, 2017). We expect immoral industries to pay a positive wage premium, a compensating differential for the immoral nature of the work.

Set of immoral industries: We select the set of immoral industries based on the industry ratings. Most participants that rated the immorality of the industries agreed that it is immoral to work in the following four industries: manufacture of weapons and ammunitions, manufacture of tobacco products, wholesale of tobacco products and gambling and betting activities (see Table B1). These four industries are also typically considered to be "sin industries" in the literature on sin stocks. We focus on these four industries.[10] We use the entire dataset as control industries.

Results: Table B6 reports regressions of the natural logarithm of real gross hourly wages on the dummies for working in each of the four immoral industries, along with several additional control variables. All four immoral industries pay substantial wage premiums, in line with an immorality premium for immoral work. According to Model 3, individuals working in the immoral industries have (geometric) mean hourly earnings of between 12 percent and 29 percent higher than people working in other industries.

[10] In Section 4, we select two other industries, monetary intermediations and credit granting, that might potentially be perceived as immoral. As there was disagreement on the immorality of these two industries among survey respondents (see Table B1), we do not include them in the set of immoral industries. Note, however, that both industries pay substantial wage premiums. If we add dummies for working in these industries to Model 3, coefficients are 0.263 (t=11.12, p<0.001) for monetary intermediation and 0.282 (t=20.51, p<0.001) for credit granting.

Table B1 Set of industries and industry ratings, initial set of industries

	Industry	Perceived immorality
Immoral industries	Manufacture of weapons and ammunitions	0.71
	Manufacture of tobacco products	0.47
	Wholesale of tobacco products	0.44
	Monetary intermediations	0.11
	Credit granting	0.15
	Gambling and betting activities	0.42
Control industries	Non-life insurance	-0.13
	Organization and operation of sport facilities	-0.49
	Processing of tea and coffee	-0.11
	Manufacture of electronic components	-0.01
	Wholesale of perfume and cosmetics	0.12
Other Industries	Manufacture of paper and paperboard	-0.06
	Construction of buildings	-0.28
	Maintenance and repair of motor vehicles	-0.28
	Wholesale of clothing and footwear	0.10
	Wholesale of watches and jewelry	0.04
	Hotels and similar accommodation	-0.34
	Restaurants and mobile food activities	-0.33
	General public administration activities	-0.41
	Fitness facilities	-0.35

Table B2: Relationship between wages and perceived industry immorality (ratings B.1)

Dependent variable: ln of real gross hourly wage (in 2010 CHF)			
	(1)	(2)	(3)
Perceived industry immorality	0.201***	0.159***	0.138***
(standardized)	(2.95)	(4.15)	(3.85)
Age		0.005***	0.005***
		(3.32)	(3.64)
Male		0.204***	0.203***
		(6.60)	(6.80)
Married		0.033*	0.037**
		(1.88)	(2.13)
Education high		0.469***	0.442***
		(8.38)	(8.64)
Education middle		0 .177***	0.170***
		(4.53)	(4.77)
Swiss		0.028	0.036*
		(1.15)	(1.97)
Experience		0.005**	0.005***
		(2.81)	(3.17)
Full-time equivalent		-0.037	-0.038
		(-0.51)	(-0.54)
Managerial duties		0.059	0.065
		(0.71)	(0.82)
Industry sales		0.027	0.034
		(0.94)	(1.25)
Industry size (employees)		0.001***	0.001***
		(3.77)	(3.69)
Constant	3.759***	2.930***	
	(70.71)	(26.43)	
N	32,638	32,638	32,638
Adjusted R^2	0.140	0.379	0.397
Year FE	No	No	Yes
Region FE	No	No	Yes

Source: Weighed data from the SLFS, years 2010-2016 (wage and demographics), STATENT, years 2011-2016 (industry size, industry sales), Value Added Tax Statistics, years 2010-2016 (industry sales) and our own survey (perceived industry immorality).

*Notes: Perceived immorality is standardized. Control variables: Male in {0, 1}, Married in {0, 1}, Education high: higher vocational education and training or university/college, Education middle: apprenticeship, full-time vocational school, matura or pedagogical training, Education low (reference category): compulsory schooling or pre-vocational education, Swiss in {0, 1}, Experience = number of years in the firm, Full-time equivalent = (working hours /42), set to 1 for working hours >= 42, managerial duties in {0, 1}, Industry size = number employees in this industry / 1000 (2010 data is not available, we substitute it with 2011 data), Industry sales = Industry sales/number employees in this industry. Model (3) controls for company region fixed effects (26 Swiss cantons) and year fixed effects (2010-2016). Standard errors clustered at the industry level, t-statistics in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

Table B3: Set of industries and industry ratings, second set of industries

	Industry	Perc. Imm. B2	Perc. Imm. B3
Immoral industries	Manufacture of weapons and ammunition	0.60	0.45
	Wholesale of tobacco products	0.54	0.39
	Processing and preserving of meat (except poultry meat)	0.24	-0.19
	Credit granting	0.23	0.15
	Processing and preserving of poultry meat	0.24	-0.16
Moral industries	Social work activities without accommodation for the elderly and disabled	-0.46	-0.70
	Residential care activities for the elderly and disabled	-0.70	-0.68
	Fire service activities	-0.73	-0.73
	Primary education	-0.78	-0.65
	Hospital activities	-0.64	-0.70
Other Industries	Manufacture of prepared meals and dishes	-0.03	0.01
	Wholesale of office machinery and equipment, except computers and computer peripheral equipment	-0.09	-0.23
	Publishing of newspapers	-0.25	-0.27
	Monetary intermediation (cantonal banks, commercial banks, stock exchange banks, private bankers; banks with a special field of business; regional banks; Raiffeisen banks; Foreign-controlled banks)	0.14	-0.01
	Passenger rail transport	-0.36	-0.57
	Support activities for crop production (preparation of fields; establishing a crop; treatment of crops; crop spraying; trimming of fruit trees and vines; transplanting of rice; thinning of beets; harvesting; pest control; provision of agricultural machinery with operators and crew)	-0.26	-0.45
	Driving school	-0.21	-0.37
	Security and commodity contracts brokerage	-0.01	-0.03
	Printing of newspapers	-0.2	-0.2
	Growing of other non-perennial crops (growing of swedes, mangolds, fodder roots, clover, alfalfa, sainfoin, fodder maize and other grasses; buckwheat; potted and bedding plants; beet seeds (excluding sugar beet seeds); seeds of forage plants and flower seeds; forage kale and similar forage products; production of cut flowers)	-0.26	-0.42
	Plant propagation	-0.24	-0.52
	Packaging activities (bottling of liquids; packaging of solids; security packaging of pharmaceutical preparations; labelling, stamping and imprinting; parcel-packing and gift-wrapping)	-0.03	-0.20
	Manufacture of fasteners and screw machine products	-0.14	-0.35
	Construction of residential and non-residential buildings	-0.15	-0.36
	Manufacture of machinery for textile, apparel and leather production	0.06	-0.28
	General medical practice activities	-0.58	-0.68
	Activities of holding companies	0.11	0.07

See next page for the rest of the table.

	Industry	Perc. Imm. B2	Perc. Imm. B3
Other Industries	Retail sale of electrical household appliances in specialised stores	-0.19	-0.29
	Wholesale trade of motor vehicle parts and accessories	-0.01	-0.20
	Wholesale of flowers and plants	-0.24	-0.26
	Dispensing chemist in specialised stores	-0.19	-0.44
	Administration of financial markets	0.10	0.03
	Manufacture of other food products (soups and broths; artificial honey and caramel; perishable prepared foods; food supplements; yeast; extracts and juices; non-dairy milk and cheese substitutes; egg products; artificial concentrates)	-0.16	-0.19
	Other personal service activities (astrological and spiritualists' activities; social activities; pet care services; genealogical organisations; tattooing and piercing studios; shoe shiners; porters; valet car parkers; concession operation of coin-operated personal service machines)	-0.18	-0.21
	Manufacture of computers and peripheral equipment	-0.07	-0.28
	Joinery installation	-0.19	-0.43
	Wholesale of beverages	-0.08	-0.27
	Camping grounds, recreational vehicle parks and trailer parks	-0.28	-0.41
	Manufacture of optical instruments and photographic equipment	-0.21	-0.37
	Renting and leasing of cars and light motor vehicles	-0.04	-0.05
	Wholesale of other machinery and equipment (transport equipment except motor vehicles; production-line robots; wires and switches; other electrical material; machinery for use in trade, navigation and industry [except mining, construction, civil engineering and textile industry]; measuring instruments and equipment)	-0.10	-0.19
	Non-specialised wholesale trade	-0.09	-0.13
	Mixed Farming	-0.17	-0.50
	Manufacture of electric domestic appliances	-0.07	-0.34
	Life insurance	0.06	-0.07
	Manufacture and processing of other glass, including technical glassware (laboratory, hygienic or pharmaceutical glassware; clock or watch glasses, optical glass and optical elements not optically worked; glassware used in imitation jewellery; glass insulators and glass insulating fittings; glass envelopes for lamps; glass figurines; glass paving blocks; glass in rods or tubes)	-0.17	-0.34
	Wholesale of pharmaceutical goods	0.01	-0.18
	Taxi operation	-0.14	-0.26
	Child day-care activities	-0.69	-0.56
	Manufacture of rusks and biscuits; manufacture of preserved pastry goods and cakes	-0.11	-0.32

Notes: Immoral (Moral) industries are the industries that the research assistants selected as the most immoral (moral) industries. Other industries are 40 randomly selected industries that have at least 50 observations in the Swiss Labor Force Survey.

Table B4: Relationship between wages and industry immorality (ratings B.2 and B.3)

Dependent variable: ln of real gross hourly wage (in 2010 CHF)						
	(1)	(2)	(3)	(4)	(5)	(6)
Perceived industry Immorality (Standardized)	0.112**	0.083***	0.072***	0.121***	0.101***	0.087***
	(2.34)	(2.94)	(2.77)	(2.78)	(5.49)	(4.87)
Age		0.008***	0.007***		0.008***	0.007***
		(7.20)	(6.94)		(7.14)	(6.90)
Male		0.152***	0.155***		0.156***	0.159***
		(7.50)	(7.62)		(6.58)	(6.67)
Married		0.030	0.032		0.029	0.031
		(1.26)	(1.33)		(1.21)	(1.29)
Education high		0.602***	0.580***		0.592***	0.572***
		(13.69)	(12.91)		(13.83)	(12.95)
Education middle		0.297***	0.288***		0.290***	0.282***
		(6.93)	(6.60)		(6.96)	(6.61)
Swiss		-0.011	-0.002		-0.013	-0.004
		(-0.63)	(-0.11)		(-0.74)	(-0.26)
Experience		0.004**	0.005***		0.004**	0.005***
		(2.49)	(2.77)		(2.59)	(2.84)
Full-time equivalent		-0.166**	-0.168**		-0.166**	-0.168**
		(-2.35)	(-2.44)		(-2.20)	(-2.28)
Managerial duties		0.063	0.066		0.062	0.066
		(0.85)	(0.93)		(0.84)	(0.92)
Industry sales		0.020	0.021		0.010	0.012
		(1.17)	(1.22)		(0.76)	(0.89)
Industry size (employees)		0.001***	0.001***		0.002***	0.001***
		(3.80)	(4.08)		(4.86)	(4.99)
Constant	3.737***	2.876***		3.819***	2.965***	
	(52.83)	(32.09)		(48.47)	(30.22)	
N	47,935	47,935	47,935	47,935	47,935	47,935
Adjusted R^2	0.041	0.248	0.263	0.049	0.254	0.267
Year and Region FE	No	No	Yes	No	No	Yes

Source: Weighed data from the SLFS, years 2010-2016 (wage and demographics), STATENT, years 2011-2016 (industry size, industry sales), Value Added Tax Statistics, years 2010-2016 (industry sales) and our own surveys (perceived industry immorality).

*Notes: Specification (1)-(3) use the perceptions for the second set of industries measured with a student sample; specifications (4)-(6) use the perceptions for the second set of industries measured with a sample that is representative of the Swiss population. Perceived immorality is standardized. Control variables: Male in {0, 1}, Married in {0, 1}, Education high: higher vocational education and training or university/college, Education middle: apprenticeship, full-time vocational school, matura or pedagogical training, Education low (reference category): compulsory schooling or pre-vocational education, Swiss in {0, 1}, Experience = number of years in the firm, Full-time equivalent = (working hours /42), set to 1 for working hours >= 42, managerial duties in {0, 1}, Industry size = number employees in this industry / 1000 (2010 data is not available, we substitute it with 2011 data), Industry sales = Industry sales/number employees in this industry. Model 3 and 6 control for company region fixed effects (26 Swiss cantons) and year fixed effects (2010-2016). Standard errors clustered at the industry level, t-statistics in parentheses; * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.*

Table B5: Summary Statistics representative sample

Variable	*Mean*	*SD*	*Min*	*Max*
Sociodemographics				
Age	45.6	16.0	16	79
Female	0.495	0.501	0	1
Total Monthly Household Income				
< 2’000 CHF	0.042			
2’001 – 3’000 CHF	0.021			
3’001 – 4’000 CHF	0.073			
4’001 – 5’000 CHF	0.07			
5’001 – 6’000 CHF	0.143			
6’001 – 7’000 CHF	0.087			
7’001 – 8’000 CHF	0.098			
8’001 – 9’000 CHF	0.087			
9’001 – 10’000 CHF	0.094			
10’001 – 11’000 CHF	0.066			
11’001 – 12’000 CHF	0.052			
12’001 – 13’000 CHF	0.032			
13’001 – 14’000 CHF	0.021			
14’001 – 15’000 CHF	0.028			
> 15’000 CHF	0.084			
German speaking	0.733	0.443	0	1

Table B6: Relationship between wages and industry immorality, dummy approach

Dependent variable: ln of real gross hourly wage (in 2010 CHF)			
	(1)	(2)	(3)
Manufacture weapons and ammunitions	0.345***	0.154***	0.147***
	(14.98)	(10.31)	(8.24)
Manufacture tobacco	0.285***	0.237***	0.288***
	(12.37)	(7.61)	(9.16)
Wholesale tobacco	0.402***	0.327***	0.280***
	(17.47)	(23.49)	(16.29)
Gambling and betting	0.025	0.087***	0.123***
	(1.08)	(7.26)	(9.44)
Age		0.006***	0.006***
		(12.66)	(12.58)
Male		0.194***	0.195***
		(17.26)	(17.81)
Married		0.035***	0.039***
		(4.57)	(5.00)
Education high		0.586***	0.564***
		(31.24)	(32.32)
Education middle		0.247***	0.241***
		(17.07)	(16.99)
Swiss		0.0024	0.011
		(0.20)	(1.01)
Experience		0.004***	0.005***
		(5.75)	(6.24)
Full-time equivalent		-0.076***	-0.078***
		(-2.59)	(-2.73)
Managerial duties		-0.025	-0.020
		(-0.91)	(-0.74)
Industry sales		0.006**	0.005**
		(2.15)	(2.04)
Industry size (employees)		0.0004	0.0004*
		(1.59)	(1.83)
Constant	3.561***	2.832***	
	(154.57)	(79.63)	
N	239,313	236,625	236,625
Adjusted R^2	0.001	0.206	0.221
Year and Region FE	No	No	Yes

Source: Weighed data from the SLFS, years 2010-2016 (wage and demographics), STATENT, years 2011-2016 (industry size, industry sales) and Value Added Tax Statistics, years 2010-2016 (industry sales).

*Notes: Manufacture weapons and ammunitions, Manufacture tobacco, Wholesale tobacco and Gambling and betting are binary variables where 1 means that the individual works in the respective industry. Control variables: Male in {0, 1}, Married in {0, 1}, Education high: higher vocational education and training or university/college, Education middle: apprenticeship, full-time vocational school, matura or pedagogical training, Education low (reference category): compulsory schooling or pre-vocational education, Swiss in {0, 1}, Experience = number of years in the firm, Full-time equivalent = (working hours /42), set to 1 for working hours >= 42, managerial duties in {0, 1}, Industry size = number employees in this industry / 1000 (2010 data is not available, we substitute it with 2011 data), Industry sales = Industry sales/number employees in this industry. Model 3 controls for company region fixed effects (26 Swiss cantons) and year fixed effects (2010-2016). Standard errors clustered at the industry level, t-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*

Appendix C – A simple model of heterogeneous moral concern in labor markets

In this section, we introduce a simple stylized model of labor markets with varying degrees of perceived job immorality and heterogeneity in concern for moral behavior among workers.

C.1 Basic model

We examine a single labor market for a job, $j \in J$, which might involve doing immoral work. Firms decide whether to hire a worker to do j at the market wage, w, and workers decide whether to accept work j for the market wage. Workers differ in their concerns for morality. We investigate how the equilibrium wage and selection in this labor market change with the immorality of j. Our framework is a simplification of the theoretical literature on compensating wage differentials (see, e.g., Rosen, 1986).[11] We do not seek to expand this literature, but rather to apply it to a context in which the relevant job dimension is immorality.

The immorality of j is measured by a function $I: J \to [0, \infty)$, where $I(j') > I(j)$ means that job j' is more immoral than job j, and $I(j) = 0$ means that j involves no immoral acts. The set of immoral jobs is $J^{IM} = \{j \in J: I(j) > 0\}$.

Labor demand is represented by an interval of firms, $k \in [0,1]$. Firms' behavior is given by the labor demand function, $D: \mathbb{R} \times J \to [0,1]$, with $lim_{w\to\infty} D(w, j) = 0, D(w, j) = 1$ for $w \leq 0$, D continuous in w and D strictly decreasing in w on $[0, \infty)$. In addition, we assume that an increase in the immorality of the job does not decrease the profitability of employing labor, that is, $I(j') > I(j)$ implies $D(w, j') \geq D(w, j)$ for all w.[12]

Labor supply consists of an interval of workers, $i \in [0,1]$. Each worker has reservation utility $\underline{u} \geq 0$, and the utility of accepting job j for a worker of type i is given by:[13]

[11] Unlike most models of compensating wage differentials, we do not have multiple labor markets, rather one and a fixed outside option. In our first laboratory experiment, we also assign subjects to one labor market. This abstraction simplifies both the theory and the experiment. However, we show in Appendix C.2 that the model allows for an interpretation with two types of jobs, an immoral job and a neutral job. Our results also apply to such a context. In our second laboratory experiment, we consider such a setting with two jobs.

[12] In our experiment, we vary the immorality of the job, but fix labor demand, that is, $D(w, j') = D(w, j)$ for all w and all $j, j' \in J$. If an increase in immorality were to decrease profitability, there would be no incentives for firms to operate in immoral industries. Heidhues, Kőszegi and Murooka (2017) provide a basis for why deceptively marketed socially harmful products may be more profitable in the presence of naïve consumers. In Appendix C.2, we provide a behavioral foundation for the labor demand.

[13] Models about "mission-oriented" employees commonly assume very similar additive utility functions (e.g. Cassar and Meier, 2018), with the main difference that $-I(j) * \theta_i$ is replaced by a positive term, the "meaningfulness of work" multiplied by how much the individual cares about meaning.

$$u_i^{accept}(j,w) = w - c - \theta_i * I(j),$$

where $c \geq 0$ is the worker's cost of effort, which is independent of j. The parameter $\theta_i \geq 0$ measures the worker's aversion to immoral work and is distributed according to a cumulative density function $F \in \mathcal{F}_\theta$. The set $\mathcal{F}_\theta$ consists of all density functions F that are continuous, strictly increasing on $[0,\infty)$, and with $F(0) = 0$, meaning that no worker likes immoral work.[14] The indirect utility of a worker of type i is then given by $v_i(j,w) = max\{\underline{u}, u_i^{accept}(j,w)\}$. Workers' behavior determines the labor supply, $S: \mathbb{R} \times J \rightarrow [0,1]$. If $j \in J^{IM}$, every worker with $\theta_i \leq \frac{w-\underline{u}-c}{I(j)}$ accepts the job. Labor supply is therefore $S(w,j) = F(\frac{w-\underline{u}-c}{I(j)})$.[15]

We next consider the equilibrium properties of this type of market. The equilibrium wage, $w^*(j)$, is implicitly defined by $S(w^*(j),j) - D(w^*(j),j) = 0$.[16] The following Lemma states that for every $j \in J$, $w^*(j)$ exists and is unique.

Lemma. *For all $j \in J^{IM}$, $w^*(j)$ exists, is unique and is in $(\underline{u} + c, \infty)$. For all $j \in J \setminus J^{IM}$, $w^*(j) = \underline{u} + c$.*

Proof. Suppose $j \in J^{IM}$. *Existence*: Define $f(w,j) = S(w,j) - D(w,j)$. Note that $f(\underline{u} + c, j) = 0 - (+) < 0$, $lim_{w\rightarrow\infty} f(w,j) = 1 - 0 = 1$ and $f(w,j)$ is continuous in w. By the intermediate value theorem there exists $w^*(j) \in (\underline{u} + c, \infty)$ such that $f(w^*(j),j) = 0$.

Uniqueness: Follows from $f(w,j)$ being strictly increasing in w on $[0,\infty)$. Suppose $j \in J \setminus J^{IM}$. Then, $S(w,j) = \begin{cases} 0, & w < \underline{u} + c \\ [0,1], & w = \underline{u} + c \\ 1, & w > \underline{u} + c \end{cases}$. Note that for any $w < \underline{u} + c$, we have $D(w,j) > 0$ but $S(w,j) = 0$, and for any $w > \underline{u} + c$ we have $D(w,j) < 1$, but $S(w,j) = 1$. For $w = \underline{u} + c$, $S(w,j) = [0,1]$ and $D(w,j) \in [0,1]$, so $D(w,j) \in S(w,j)$. ■

[14] Note that $F(0) = 0$ implies that no worker likes to do immoral jobs.

[15] The assumptions on F (together with the properties of a cdf) imply that S is continuous and strictly increasing in w on $[\underline{u} + c, \infty)$, $\lim_{w\rightarrow\infty} S(w,j) = 1$, and $S(w,j) = 0$ for all $w \leq \underline{u} + c$.

[16] Note that for $j \in J \setminus J^{IM}$, S is a correspondence. For this case, $w^*(j)$ is defined by $D(w^*(j),j) \in S(w^*(j),j)$. Moreover, $w^*(j)$ depends on F. When necessary (Corollary) we will make this explicit by writing $w^*(j,F)$ instead of $w^*(j)$.

In the following, we derive four properties of labor markets with immoral jobs. While straightforward, we use these results to make predictions for our empirical work. In particular, the first two propositions derive the primary hypotheses that we test across all of our analysis.

Proposition 1 shows that there is an immorality premium for immoral jobs: an increase in the immorality of a job decreases supply and therefore increases the equilibrium wage.

Proposition 1. (Immorality premium) *For all $j, j' \in J$ with $I(j) < I(j')$, $w^*(j) < w^*(j')$.*

Proof. $w^*(j) < w^*(j')$: Suppose $I(j) = 0$. Then, $w^*(j) = \underline{u} + c$ and $w^*(j') > \underline{u} + c$ (see Lemma). Suppose $I(j) > 0$. Suppose that $w^*(j) \geq w^*(j')$. Then $S(w^*(j), j) > S(w^*(j'), j) > S(w^*(j'), j')$ because $w^*(j') > \underline{u} + c$ (see Lemma) and S is strictly increasing in w and decreasing in $I(j)$ on $[\underline{u} + c, \infty)$. Moreover, $D(w^*(j), j) \leq D(w^*(j'), j) \leq D(w^*(j'), j')$ because $-D$ is strictly increasing in w on $[0, \infty)$ and $\mathrm{I}(j) < I(j')$ and, therefore, $D(w, j') \geq D(w, j)$ for all w. So $S(w^*(j), j) - S(w^*(j'), j') + D(w^*(j'), j') - D(w^*(j), j) > 0$, a contradiction to the definition of $w^*(j)$ and $w^*(j')$. ■

The following Corollary further shows that this wage premium will be insignificant if workers do not sufficiently care about morality.

Corollary. *For all $j, j' \in J$ with $I(j) < I(j')$ and $\varepsilon > 0$, there exists $G \in \mathcal{F}_\theta$ such that $w^*(j', G) - w^*(j, G) \leq \varepsilon$.*

Proof. Suppose that there exists a $G \in \mathcal{F}_\theta$ such that

$$G\left(\frac{\varepsilon}{I(j')}\right) = D(\underline{\mathrm{u}} + \mathrm{c} + \varepsilon, j').$$

Then, $\mathrm{w}^*(\mathrm{j}', \mathrm{G}) = \underline{\mathrm{u}} + \mathrm{c} + \varepsilon$. The Lemma and Proposition 1 then imply that $\mathrm{w}^*(\mathrm{j}, \mathrm{G}) \in [\underline{\mathrm{u}} + \mathrm{c}, \underline{\mathrm{u}} + \mathrm{c} + \varepsilon)$, and, as a result, $w^*(j', G) - w^*(j, G) \leq \varepsilon$.

To proof that such a $G \in \mathcal{F}_\theta$ exist, take any $H \in \mathcal{F}_\theta$ and construct G as follows:

$$G(\mathrm{x}) = \begin{cases} 0 & if\ \mathrm{x} < 0 \\ x\frac{I(j')}{\varepsilon}D(\underline{\mathrm{u}} + \mathrm{c} + \varepsilon, j') & if\ x \in [0, \frac{\varepsilon}{I(j')}] \\ D(\underline{\mathrm{u}} + \mathrm{c} + \varepsilon, j') + \left(1 - D(\underline{\mathrm{u}} + \mathrm{c} + \varepsilon, j')\right)H(x - \frac{\varepsilon}{I(j')}) & if\ \mathrm{x} > \frac{\varepsilon}{I(j')} \end{cases}$$

The assumptions on D imply that $D(\underline{\mathrm{u}} + \mathrm{c} + \varepsilon, \mathrm{j}') \in (0,1)$. Note that G is continuous, strictly increasing on $[0, \infty)$, and with $G(0) = 0$. Therefore $G \in \mathcal{F}_\theta$. ∎

Formally, the Corollary shows that there are distributions of moral types such that the wage differentials are arbitrary small.

Second, the types that care least about the immorality of a job ($\theta_i \leq \frac{w^*(j) - \underline{u} - c}{I(j)}$), sort into accepting immoral jobs, while those more concerned with morality ($\theta_i > \frac{w^*(j) - \underline{u} - c}{I(j)}$), refuse to do the job for the equilibrium wage.[17] This is formally shown in Proposition 2.

Proposition 2. (Sorting) *For all $j \in J^{IM}$, worker i is hired iff* $\theta_i \leq \frac{w^*(j) - \underline{u} - c}{I(j)} \equiv \underline{\theta}(j) > 0$.

Proof. A worker accepts job j iff $u_i^{accept} = w^*(j) - c - \theta_i * I(j) \geq \underline{u} \Leftrightarrow \theta_i \leq \frac{w^*(j) - c - \underline{u}}{I(j)}$.

$\underline{\theta}(j) > 0$: Follows from $w^*(j) > \underline{u} + c$ (see Lemma). ∎

Proposition 2 is critical to the notion that immorality wage premiums are driven by those who find immoral work most distasteful opting out of such jobs.

C.2 Alternative model interpretation for a context with 2 jobs

The results in Section 3 and Appendix C.1 also apply for a context with 2 jobs, a neutral job j^N ($I(j^N) = 0$) and an immoral job $j^{IM} \in J^{IM}$ ($I(j^{IM}) > 0$). In the following, we show that, under some assumptions, labor demand and labor supply correspond to their counterparts in Appendix C.1. Therefore, all results derived in Appendix C.1 also hold in the context with 2 jobs.

[17] This perfect sorting according to θ is an extreme case. Heterogeneity in the costs of effort, reservation utility or productivity implies partial sorting according to θ (see Garen (1988) and Hwang, Reed and Hubbard (1992) for related examples). We incorporate only one dimension of heterogeneity in the model for simplicity.

Labor supply: Labor supply consists of an interval of workers, $i \in [0,1]$. As in Appendix C.1, we assume that the utility of accepting job j of a worker of type i is given by:

$$u_i\big(j, w(j)\big) = w(j) - c - \theta_i * I(j),$$

where the parameter θ_i is distributed according to a distribution with cdf, $F \in \mathcal{F}_\theta$. The set $\mathcal{F}_\theta$ consists of all density functions F that are continuous, strictly increasing on $[0, \infty)$, and with $F(0) = 0$. Workers choose between the neutral and the immoral job. Note that every worker with $\theta_i \leq \frac{w(j^{IM}) - w(j^N)}{I(j^{IM})}$ chooses the immoral job. The labor supply for the immoral job is then given by $S(w, j^{IM}) = F(\frac{w}{I(j^{IM})})$, where $w = w(j^{IM}) - w(j^N)$ is the immorality premium. Note that the labor supply for the immoral job corresponds to the labor supply in Appendix C.1 with $\underline{u} = c = 0$.

Labor demand: Labor demand consists of an interval of firms, $k \in [0,1]$. Each firm can either produce a neutral product or an immoral product. Firms that produce immoral products offer immoral jobs; firms that produce neutral products offer neutral jobs. Firms' profits are:

$$\pi_k(j, w) = a_k(j) - w(j),$$

where $a_k(j)$ measures firm k's earnings when producing good j. Firms choose to produce the immoral product if $\Delta a_k(j^{IM}) = a_k(j^{IM}) - a_k(j^N) \geq w$. This term, $\Delta a_k(j^{IM})$, is distributed according to a distribution with cdf $G_{j^{IM}}$. An increase in immorality of the job does not decrease firms earnings,[18] that is, i) $G_j(0) = 0$ for all $j \in J^{IM}$, and ii) for all $j, j' \in J^{IM}$ with $I(j') > I(j)$, $G_{j'}(x) \leq G_j(x)$ for all $x \in \mathbb{R}$. In addition, $G_{j^{IM}}$ is continuous and strictly increasing on $[0, \infty)$. The labor demand for the immoral job is then given by $D(w, j^{IM}) = 1 - G_{j^{IM}}(w)$. Note that D is continuous in w, strictly decreasing in w on $[0, \infty)$, with $lim_{w \to \infty} D(w, j^{IM}) = 0$ and $D(w, j^{IM}) = 1$ for $w \leq 0$. In addition, $I(j') > I(j)$ implies $D(w, j') \geq D(w, j)$ for all $w \in \mathbb{R}$. Note that the labor demand satisfies all assumptions made for the labor demand in Appendix C.1.

The equilibrium wage, $w^*(j^{IM})$, is implicitly defined by $S(w^*(j^{IM}), j^{IM}) - D(w^*(j^{IM}), j^{IM}) = 0$.[19] As both labor demand and labor supply satisfy all assumptions made in

[18] One interpretation is, for example, that $I(j)$ measures negative externalities in production. Avoiding these externalities is costly; decreasing the immorality therefore increases production costs (see also Rosen, 1986).

[19] Note that market clearance in the immoral job market implies market clearance in the neutral job market, $\big(1 - S(w^*(j^{IM}), j^{IM})\big) - \big(1 - D(w^*(j^{IM}), j^{IM})\big) = 0$.

Appendix C.1, the Lemma, Corollary and Proposition 1 and 2 apply (with $j \in J^{IM}$). In particular, $w^*(j^{IM})$ is strictly positive (Lemma), so there is an immorality premium, and this immorality premium is increasing in the immorality of j^{IM}, $I(j^{IM})$ (Proposition 1). The immoral types sort into accepting the immoral jobs, while the moral types sort into accepting the neutral jobs (Proposition 2).

C.3 Extended model: Workers can impact job immorality

In Appendix C.1, we assume that workers cannot impact job immorality, $I(j)$. Suppose instead that a hired worker can chose to reduce job immorality (e.g., reduce harm) by an amount $e < I(j)$ at a cost c^{RH}. That is, the worker either takes the action a_i to reduce job immorality ($a_i = 1$) or to not reduce job immorality ($a_i = 0$). Otherwise, we use the same model and assumptions as in Appendix C.1. A worker of type θ_i then accepts job j if the utility from doing so is higher than that of an outside option, or

$$u_i^{accept}(j, w|a_i = 0) = w - c - \theta_i I(j) \geq \underline{u},$$

$$u_i^{accept}(j, w|a_i = 1) = w - c - \theta_i (I(j) - e) - c^{RH} \geq \underline{u},$$

where $c \geq 0$ is the worker’s cost of effort and $\underline{u} \geq 0$ is the workers’ reservation utility. As before, θ_i captures an aversion to personally implementing immoral work (as with “impure” motives such as warm glow or image concerns). Note that the utility of accepting the job depends on the worker’s decision whether to reduce job immorality. The worker reduces job immorality if $u_i^{accept}(j, w|a_i = 1) > u_i^{accept}(j, w|a_i = 0)$ or $\theta_i > c^{RH}/e$.

If $j \in J^{IM}$, every worker with $\theta_i \leq \max\left(\frac{w-c-\underline{u}}{I(j)}, \frac{w-c-\underline{u}-c^{RH}}{I(j)-e}\right)$ accepts the job (see proof of Proposition 4). Labor supply is therefore $S(w, j) = F\left(\max\left(\frac{w-c-\underline{u}}{I(j)}, \frac{w-c-\underline{u}-c^{RH}}{I(j)-e}\right)\right)$. Proposition 3 shows that there exists a unique market wage $w^*(j)$ that clears the market and that there is an immorality premium for immoral jobs.

Moreover, Proposition 4 shows that the model predicts that the types that care least about the immorality of a job ($\theta_i \leq \underline{\theta}(j)$), sort into accepting immoral jobs, while those more concerned with morality ($\theta_i > \underline{\theta}(j)$), refuse to do the job for the equilibrium wage. Hence, the model predicts that the aversion to implementing immoral work cause prosocial individuals to avoid such work, even when they can do "good" by entering such industries.[20]

Next, we show that these predictions stand in stark contrast to the sorting of individuals into "moral" work. In the case of such work, the model predicts that individuals who prioritize morality will be more likely to sort into moral work, and their presence will lead to an increase in overall job morality. To show this, consider a job $j \in J^M$ that involves doing moral work. A hired worker can increase job morality, $M(j) > 0$, by an amount e (e.g., increase social benefit) at a cost of c^{IB}.

Note that when $c^{IB} = c^{RH}$, from a consequentialist standpoint, immoral and moral work are equivalent, as both allow a worker to benefit society by an amount e at a cost $c^{IB} = c^{RH}$. Hence, a model that assumes workers to be motivated only by consequentialist (or, "pure") social motives would predict morally motivated workers to find entering moral and immoral work equally attractive.

However, instead of assuming such consequentialist social motives, our model focuses on workers that are influenced by an aversion to personally implementing immoral work and a corresponding preference for performing moral work. A worker of type θ_i then accepts job j if the utility from doing so is higher than that of an outside option, or

$$u_i^{accept}(j, w | a_i = 0) = w - c + \theta_i M(j) \geq \underline{u},$$

$$u_i^{accept}(j, w | a_i = 1) = w - c + \theta_i (M(j) + e) - c^{IB} \geq \underline{u}$$

where θ_i measures a workers concerned with acting morally, $c \geq 0$ is the worker's cost of effort and $\underline{u} \geq 0$ is the workers' reservation utility and $a_i = 1$ means that the worker increases job morality. As for the case of immoral work, a worker choses the moral action $a_i = 1$ if $\theta_i > c^{IB}/e$.

[20] Note that a model that assumes workers to be motivated only by consequentialist concerns to reduce harm, ϑ_i, would make the opposite prediction. Such a worker would accept job j if the utility from doing so is higher than that of an outside option, or

$$u_i^{accept}(j, w | a_i = 0) = w - c \geq \underline{u},$$

$$u_i^{accept}(j, w | a_i = 1) = w - c + \vartheta_i e - c^{RH} \geq \underline{u}$$

and would reduce job immorality if $\vartheta_i > c^{RH}/e$. Note that the immoral job is *more attractive for moral workers* with $\vartheta_i > c^{RH}/e$, (who receive utility $u_i^{accept} = u_i^{accept}(j, w | a_i = 1) > w - c$) than for workers that are less concerned with morality with $\vartheta_i < c^{RH}/e$ (who receive utility $u_i^{accept} = u_i^{accept}(j, w | a_i = 0) = w - c$). Hence, such a model would predict *moral* workers to sort into immoral work.

That is, the same condition applies to a hired worker's decision to increase job morality as to the decision to decrease job immorality.

Every worker with $\theta_i \geq \min\left(\frac{\underline{u}+c-w}{M(j)}, \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}\right)$ accepts the job (see proof of Proposition 4). Labor supply is therefore $S(w,j) = 1 - F\left(\min\left(\frac{\underline{u}+c-w}{M(j)}, \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}\right)\right)$.[21] Proposition 3 states that for every j with $M(j) > 0$, $w^*(j)$ exists, is unique and, importantly, is below $\underline{u} + c$. Note that this result implies the existence of a "morality wage discount" for moral work.

Proposition 4 then show that, in contrast to the case of immoral work, the types that care most about the morality of a job ($\theta_i \geq \bar{\theta}(j)$), sort into accepting moral jobs, while those less concerned with morality ($\theta_i < \bar{\theta}(j)$), refuse to do the job for the equilibrium wage.

Proposition 5 compares moral behavior ($a_i = 1$) in the moral work and in immoral work conditions. Note that we refer to the share of hired works that choose the moral action $a_i = 1$ (that is, they either increase job morality for j ∈ J^M, or reduce job immorality for immoral j ∈ J^{IM}) as $E(a,j)$. A first key implication of the proposition is that the moral action is chosen more often for moral work than for immoral work ($E(a,j^M) \geq E(a,j^{IM})$) if the costs of reducing job immorality is the same as the costs of increasing job morality (that is, $c^{RH} = c^{IB}$). This effect is driven by differential sorting (Proposition 4).

In our experiment, we equalize the *monetary* costs of reducing job immorality and of increasing job morality, with the goal of setting $c^{RH} = c^{IB}$. Hence, we predicted to find $E(a,j^M) \geq E(a,j^{IM})$. However, while we found evidence supporting Proposition 3 and 6, we also found that $E(a,j^M) < E(a,j^{IM})$. As we discuss in the paper, a plausible explanation for this finding is that there are non-monetary costs of choosing the moral action that are a decreasing function of income. The higher market wages for immoral work (which is predicted by the theory and supported by the data) would then result in $c^{RH} < c^{IB}$. Moreover, the existence of moral licensing and moral cleansing effects would increase c^{IB} and reduce c^{RH}, respectively, thereby further increasing the difference between c^{IB} and c^{RH}. After collecting the data, we therefore generalized Proposition 5 to allow for $c^{RH} < c^{IB}$. As the propositions shows, allowing the costs

[21] Note that the assumptions on F (together with the properties of a cdf and the min function) imply that S is continuous and strictly increasing in w on $(-\infty, \underline{u} + c]$, strictly increasing in $M(j)$ on $(-\infty, \underline{u} + c]$, $\lim_{w \to -\infty} S(w,j) = 0$, and $S(w,j) = 1$ for all $w \geq \underline{u} + c$.

of the moral action to differ between moral and immoral work can indeed capture the finding that $E(a,j^M) < E(a,j^{IM})$. Note that neither Proposition 3 nor Proposition 4 requires $c^{RH} = c^{IB}$, and, hence, our model can also explain the immorality wage premium (Proposition 3) and differential sorting (Proposition 4).

Propositions 4 and 5 illustrate potentially important contrasts between moral and immoral work. The model proposes that difference between moral and immoral work cannot simply be described by "moral types" sorting into moral work to do "good" and sorting out of immoral work to reduce the amount of "bad" that takes place. Instead, the model proposes that moral types have a strong aversion to personally implementing immoral work, and, as a result, they do not accept immoral work even if accepting such work would allow them to reduce how much "bad" takes place. Such differential sorting than negatively impacts moral behavior at the workplace for moral work. In addition, there are other important differences between moral and immoral work (income, opportunity for moral licensing, need for moral cleansing) that produce further differences in moral behavior at the workplace, and can even dominate the negative effect of sorting.

Proposition 3. (Wage differences) *For all $j \in J^{IM}$, $w^*(j)$ exists, is unique and is in $(\underline{u} + c, \infty)$. For all $j \in J^M$, $w^*(j)$ exists, is unique and is in $(-\infty, \underline{u} + c)$.*

Proof. For $j \in J^{IM}$: The new labor supply satisfies all assumptions made in Appendix C.1 to prove the Lemma.[22] The lemma shows that, *for all $j \in J^{IM}$, $w^*(j)$ exists, is unique and is in* $(\underline{u} + c, \infty)$. For $j \in J^M$: *Existence*: Define $f(w,j) = S(w,j) - D(w,j)$. Note that i) $f(\underline{u} + c, j) = 1 - z > 0$ (with $z < 1$ because $D(0,j) = 1$ and D strictly decreasing in w on $[0, \infty)$), ii) $lim_{w \to -\infty} f(w,j) = 0 - 1 = -1$ and iii) $f(w,j)$ is continuous in w. By the intermediate value theorem there exists $w^*(j) \in (-\infty, \underline{u} + c)$ such that $f(w^*(j), j) = 0$. *Uniqueness*: Follows from $f(w,j)$ being strictly increasing in w on $[0, \infty)$. ■

[22] The assumptions on F (together with the properties of a cdf and the max function) imply that S is continuous and strictly increasing in w on $[\underline{u} + c, \infty)$, strictly decreasing in $I(j)$ on $[\underline{u} + c, \infty)$, $\lim_{w \to \infty} S(w,j) = 1$, and $S(w,j) = 0$ for all $w \leq \underline{u} + c$.

Proposition 4. (Differential sorting) *For all* $j \in J^{IM}$*, worker* i *is hired iff* $\theta_i \leq \max\left(\frac{w^*(j)\ -c-\underline{u}}{I(j)}, \frac{w^*(j)\ -c-\underline{u}\ -c^{RH}}{I(j)-e}\right) \equiv \underline{\theta}(j) > 0$. *For all* $j \in J^{M}$*, worker* i *is hired iff* $\theta_i \geq \min\left(\frac{\underline{u}+c-w^*(j)}{M(j)}, \frac{\underline{u}+c+c^{IB}-w^*(j)}{M(j)+e}\right) \equiv \bar{\theta}(j) > 0$.

Proof. For $j \in J^{IM}$: First note that workers accept the job if i) $\theta_i \leq \min\ (\frac{w-c-\underline{u}}{I(j)}, \frac{c^{RH}}{e})$ or if ii) $\frac{c^{RH}}{e} < \theta_i \leq \frac{w-c-\underline{u}\ -c^{RH}}{I(j)-e}$. Next, consider two cases. First, $\frac{c^{RH}}{e} \geq \frac{w-c-\underline{u}}{I(j)}$. Then, inequality i) simplifies to $\theta_i \leq \frac{w-c-\underline{u}}{I(j)}$ and there is no θ_i that satisfies inequalities ii). Hence, in this case workers accept the job if $\theta_i \leq \frac{w-c-\underline{u}}{I(j)}$. Second, $\frac{c^{RH}}{e} < \frac{w-c-\underline{u}}{I(j)}$. Then, inequality i) simplifies to $\theta_i \leq \frac{c^{RH}}{\mathrm{e}}$ and there exist θ_i's that satisfy inequalities ii). Hence, in this case workers accept the job if $\theta_i \leq \frac{w-c-\underline{u}\ -c^{RH}}{I(j)-e}$. Finally, note that $\frac{c^{RH}}{e} \geq \frac{w-c-\underline{u}}{I(j)}$ is equivalent to $\frac{w-c-\underline{u}}{I(j)} \geq \frac{w-c-\underline{u}\ -c^{RH}}{I(j)-e}$. This allows us to combine the two cases: workers accept the job if $\theta_i \leq \max\left(\frac{w-c-\underline{u}}{I(j)}, \frac{w-c-\underline{u}\ -c^{RH}}{I(j)-e}\right)$. Finally, in equilibrium $w = w^*(j)$. $\underline{\theta}(j) > 0$: Follows from $w^*(j) > \underline{u} + c$ (see Lemma).

For $j \in J^M$: First note that workers accept the job if i) $\frac{c^{IB}}{e} > \theta_i \geq \frac{\underline{u}+c-w}{M(j)}$ or if ii) $\theta_i \geq \max\ (\frac{c^{IB}}{e}, \frac{\underline{u}+c+c^{IB}-w}{M(j)+e})$. Next, consider two cases. First, $\frac{c^{IB}}{e} \geq \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}$. Then, inequality ii) simplifies to $\theta_i \geq \frac{c^{IB}}{e}$ and there exist θ_i's that satisfy inequalities ii). Hence, in this case workers accept the job if $\theta_i \geq \frac{\underline{u}+c-w}{M(j)}$. Second, $\frac{c^{IB}}{e} < \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}$. Then, inequality ii) simplifies to $\theta_i \geq \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}$ and there is no θ_i that satisfies inequalities i). Hence, in this case workers accept the job if $\theta_i \geq \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}$. Finally, note that $\frac{c^{IB}}{e} \geq \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}$ is equivalent to $\frac{c^{IB}}{e} \geq \frac{\underline{u}+c-w}{M(j)}$, which is equivalent to $\frac{\underline{u}+c+c^{IB}-w}{M(j)+e} \geq \frac{\underline{u}+c-w}{M(j)}$. This allows us to combine the two cases: workers accept the job if $\theta_i \geq \min\left(\frac{\underline{u}+c-w}{M(j)}, \frac{\underline{u}+c+c^{IB}-w}{M(j)+e}\right)$. Finally, in equilibrium $w = w^*(j)$. $\bar{\theta}(j) > 0$: Follows from $w^*(j) < \underline{u} + c$ (see Lemma moral work). ■

Proposition 5. (Moral behavior at work) *For all* $j^{IM} \in J^{IM}$*and* $j^M \in J^M$*,*

- if $\underline{\theta}(\, j^{IM}) \leq c^{RH}/e$, then $E(a, j^M) \geq E(a, j^{IM}) = 0$.
- if $\underline{\theta}(\, j^{IM}) > c^{RH}/e$, *there exist* $\underline{c}(j^M, j^{IM}) < 0$ *such that* $E(a, j^M) > E(a, j^{IM})$ *if* $c^{RH} - c^{IB} > \underline{c}(j^M, j^{IM})$ *and* $E(a, j^M) < E(a, j^{IM})$ *if* $c^{RH} - c^{IB} < \underline{c}(j^M, j^{IM})$*.*

Proof. For $j^{IM} \in J^{IM}$, $E(a, j^{IM}) = \max\left(1 - \frac{F(c^{RH}/e)}{F(\underline{\theta}(j))}, 0\right)$. That is, if $\frac{c^{RH}}{e} \geq \underline{\theta}(j)$, none of the hired workers reduce harm. If $\frac{c^{RH}}{e} < \underline{\theta}(j)$, only the workers with $\frac{c^{RH}}{e} < \theta_i \leq \underline{\theta}(j)$ reduces job immorality. Note that $\frac{F(c^{RH}/e)}{F(\underline{\theta}(j))} > 0$ because $c^{RH}/e > 0$, $\underline{\theta}(j) > 0$, $F(0) = 0$ and F strictly increasing on $[0, \infty)$, and, hence, $E(a, j^{IM}) < 1$.

For $j^M \in J^M$, $E(a, j^M) = \min\left(\frac{1-\mathrm{F}(c^{IB}/e)}{1-\mathrm{F}(\bar{\theta}(j))}, 1\right)$. That is, if $\frac{c^{IB}}{e} \leq \bar{\theta}(j)$, all of the hired workers increase job morality. If $\frac{c^{IB}}{e} > \bar{\theta}(j)$, only the workers with $\theta_i > c^{IB}/e$ increase job morality. Note that $\frac{1-\mathrm{F}(c^{IB}/e)}{1-\mathrm{F}(\bar{\theta}(j))} > 0$ because $\mathrm{F}(c^{IB}/e) < 1$ and $\mathrm{F}\left(\bar{\theta}(j)\right) < 1$ and, hence, $E(a, j^M) > 0$.

Consider the first part of the proposition. Note that when $\underline{\theta}(\, j^{IM}) \leq c^{RH}/e$, then, $E(a, j^{IM}) = 0$ while $E(a, j^M) > 0$.

Consider the second part of the proposition. Now, $\underline{\theta}(\, j^{IM}) > c^{RH}/e$. In the following, we will keep c^{RH} fix, and then construct a $c^{IB}(j^M, j^{IM})$ such that $E(a, j^{IM}) < E(a, j^M)$ is equivalent to $c^{IB} < c^{IB}(j^M, j^{IM})$ (and, hence, when $c^{IB} > c^{IB}(j^M, j^{IM})$ then $E(a, j^{IM}) > E(a, j^M)$.) First, note that it must be that $\frac{c^{IB}(j^M, j^{IM})}{e} \geq \bar{\theta}(j)$. (Proof: If $\frac{c^{IB}(j^M, j^{IM})}{e} < \bar{\theta}(j)$, then there would exist ε such that $\frac{c^{IB}(j^M, j^{IM})+\varepsilon}{e} < \bar{\theta}(j)$, and, hence, $c^{IB} = c^{IB}(j^M, j^{IM}) + \varepsilon > c^{IB}(j^M, j^{IM})$ but $E(a, j^{IM}) \leq E(a, j^M) = 1$. Hence, $c^{IB} < c^{IB}(j^M, j^{IM})$ would not be equivalent to $E(a, j^{IM}) < E(a, j^M)$.) Hence, we have to focus on cases where i) $\underline{\theta}(\, j^{IM}) > c^{RH}/e$ and ii) $c^{IB}/e \geq \bar{\theta}(j)$ to identify $\frac{c^{IB}(j^M, j^{IM})}{e}$. Then $E(a, j^{IM}) < E(a, j^M)$ is equivalent to $g(c^{IB}) = F\left(\frac{c^{RH}}{e}\right) - F\left(\underline{\theta}\right) - \frac{F\left(\frac{c^{RH}}{e}\right) - \mathrm{F}\left(\frac{c^{IB}}{e}\right) F(\underline{\theta})}{\mathrm{F}(\bar{\theta})} < 0$. Note that $g(c^{IB})$ is a continuous function, and that $\mathrm{F}\left(\frac{c^{RH}}{e}\right) - F\left(\underline{\theta}\right) < 0$ because $\frac{c^{RH}}{e} < \underline{\theta}(j)$ and F strictly increasing. Now, note that there exist a c^{IB}, namely $c^{IB} = c^{RH}$, such that $g(c^{IB}) < 0$. In this case, $\frac{F\left(\frac{c^{RH}}{e}\right)-\mathrm{F}\left(\frac{c^{IB}}{e}\right)F(\underline{\theta})}{\mathrm{F}(\bar{\theta})}$ simplifies to $\mathrm{F}\left(\frac{c^{RH}}{e}\right)\frac{1-F(\underline{\theta})}{\mathrm{F}(\bar{\theta})} \geq 0$. At the same time, note that $c^{IB} \rightarrow \infty$ results in $g(c^{IB}) > 0$. In this case, $g(c^{IB})$ converges to $\left(F\left(\frac{c^{RH}}{e}\right) - F(\underline{\theta})\right)\left(1 - \frac{1}{\mathrm{F}(\bar{\theta})}\right) > 0$ (with $\bar{\theta} = \frac{\underline{u}+c-w^*(j)}{M(j)}$). Hence, by the intermediate value theorem there exists $c^{IB}(j^M, j^{IM}) \in (c^{RH}, \infty)$ such that $g(c^{IB}(j^M, j^{IM})) = 0$. To finish the proof, define $\underline{c}(j^M, j^{IM}) = c^{RH} - c^{IB}(j^M, j^{IM}) < 0$ and note that $E(a, j^{IM}) < E(a, j^M)$ is equivalent to $g(c^{IB}) < 0$ which is equivalent to $c^{IB} < c^{IB}(j^M, j^{IM})$ which is equivalent to $c^{RH} - c^{IB} > \underline{c}(j^M, j^{IM})$. ■

Appendix D – Details design Study 2

Laboratory sessions consisted of 24 participants. Before entering the lab, we took a portrait photograph of each subject to make labor market outcomes public. As subjects could influence the amount of a donation to the NRA and Everytown, participants read an information sheet at the beginning of the experiment about gun violence in the US, and about these two organizations.

In the following, we describe each of the choices subjects completed in the laboratory session, in detail. We also provide details on the recruitment and role of the clients.

D.1 Behavioral measure of concern for morality (θ)

As in Study 1, participants first played an incentivized game that measures their willingness to lie for personal gain while causing harm to others in a non-market environment. We modified the game such that it mimics the consequences of immoral work in Study 2.

In the game, Participant A privately observes a computerized die roll, *r*, and sends a message reporting the observed number to Participant B. Participant A may send a message, *m*, claiming that the observed number is either "1", "2," "3," "4," "5," or "6," regardless of the actual number. Participant A receives 2*.5+0.5m* CHF,[23] which means that she has an incentive to lie if *r* is less than 6. Participant B then decides whether "to follow" or "not to follow" the message sent by Participant A. If Participant B does not follow the message, he receives 0.75 CHF and the donations to NRA and Everytown are unaffected. If he follows the message and Participant A truthfully reported the observed number, Participant B earns 2.5 CHF and the donation to Everytown is increased by CHF 0.5 while the donation to NRA is decreased by CHF 0.5. However, if Participant B follows the message and Participant A lied, Participant B does not earn any money and the donation to Everytown is decreased by CHF 0.5 while the donation to NRA is increased by CHF 0.5.

Every participant plays the role of Participant A. We use the strategy method to elicit Participant A's message for every possible die roll. At the end of the experimental session, we select 1 out of 5 participants, and only the choices of these randomly selected participants are implemented. However, Participant As are paid based on their choice, independently of whether

[23] During the experiment, subjects accumulated earnings in "points," which we converted to money at the rate of 20 points = 1 CHF. We present the design and results in terms of ultimate payments in Swiss francs (CHF) to provide a clearer indication of the economic significance.

or not their choices are implemented. We then recruit other participants from the same subject pool to participate in an online experiment to play the role of Participants B.

Participants were informed that, at the end of the session and after all choices had been made, their decisions as Participant A would be publicly displayed to other participants in the session, along with their portrait photograph. This was all explained clearly and publicly at the beginning of the experiment.

D.2 Labor market experiment

In the labor market, subjects play the role of workers competing to be hired by automated firms. Before interacting in the market, subjects receive general instructions about the labor market and hear a recording of the instructions. Then, participants answer comprehension questions about the market, including how prices and quantities are determined. Only after the above instructions about market procedures, subjects receive information about the nature of the job, which varies by treatment. This ensures that subjects in both conditions interpret the market instructions similarly. Their understanding of these new instructions about the job is again tested through comprehension questions.

The jobs. There are two types of jobs in each round, *G jobs* and *neutral jobs*. The neutral jobs require moving one 2 x 2 LEGO brick from the top to the bottom of a white LEGO plate. Neutral jobs do not affect the donations made either to the NRA or to Everytown. The G jobs always consists of two elements: selecting a statement and replacing LEGO bricks. The moral nature of the G job differs across treatment conditions. The G job is "immoral" in the immoral work condition, "moral" in the moral work condition, and morally neutral in the neutral work condition. In the following, we discuss the two elements of the G job.

Selecting a statement: In the immoral and moral work condition, participants have to generate (mis)information about gun violence. A participant that is hired for a G job must select one of the following two statements on the computer screen:

1. "Gun violence is a rare cause of death for children in the U.S. Child deaths from gun violence are less frequent than deaths from car accidents and cancer."
2. "Gun violence is the leading cause of death for children in the U.S. Child deaths from gun violence are more frequent than deaths from car accidents and cancer."

The participants learn that the second statement is true (Goldstick, Cunningham and Carter, 2022). Hired workers in the immoral work condition must select the first statement, while hired workers in the moral work condition must select the second statement. We then display the selected statement to US participants that are hired as part of a separate online survey study.[24] The participants that are hired to do a G job in the neutral work condition must select one correct statement from two morally neutral statements.[25]

Replacing LEGO bricks and reallocating donations: Participants find a green LEGO plate, with twelve grey 2x2 bricks and twelve grey 2x3 bricks on their table. Adjacent to the plate, a plastic bag contains twelve red 2x2 bricks and twelve red 2x3 bricks. In all treatment conditions, hired workers must replace a grey LEGO brick on the plate with a similarly sized red brick from the bag. They can decide to either replace a 2x2 brick or a 2x3 brick.

In the immoral and moral work conditions, the bricks on the green plate represent a donation to Everytown and the NRA. In the immoral work condition, the grey bricks represent donations to Everytown while the red bricks represent donations to the NRA. Replacing a 2x3 brick moves a donation of CHF 0.6 from Everytown to the NRA. If the worker replaces a 2x2 brick instead, only a donation of CHF 0.4 is moved, but the hired worker pays a cost of CHF 0.2. In terms of the model, a hired worker can reduce the amount of "harm" by $e = 0.4$ (that is, the difference between the donations to NRA and Everytown) for a cost of $c = 0.2$.

In the moral work condition, the grey bricks represent donations to the NRA while the red bricks represent donations to Everytown. Replacing a 2x2 brick moves a donation of CHF 0.4 from the NRA to Everytown. If the worker instead replaces a 2x3 brick, a donation of CHF 0.6 is moved, but the hired worker has to pay a cost of CHF 0.2. In terms of the model, a hired worker can increase the "benefit" by $e = 0.4$ (that is, the difference between the donations to NRA and Everytown) for a cost of $c = 0.2$.

[24] We recruited these participants through Prolific for a brief survey. We tell the participants in the laboratory experiment that we will not tell the U.S. participants whether the information provided is true or false, which is correct. We see this as unproblematic for three reasons. First, prior to revealing the information, we inform U.S. participants that the information about gun violence may or may not be true. Second, each U.S. participant is paired with two participants from the laboratory experiment, one from the immoral work condition and one from the moral work condition. Consequently, U.S. participants also encounter the truthful statement. Third, at the experiment's conclusion, we offer participants a link to verify the accurate statistics.

[25] Participants are presented with two statements: 1. "Switzerland is a member of both the European Union and the Schengen Area," and 2. "Switzerland is not a member of the European Union, but it is a member of the Schengen Area." Their task is to choose the correct second statement. Importantly, this choice incurs the same effort costs as the corresponding choices in the moral and immoral work conditions.

In the neutral work condition, replacing the LEGO bricks has no consequences. To keep the instructions as similar as possible across treatment condition, a hired worker must pay CHF 0.2 to replace a 2x2 brick instead of a 2x3 brick.

If a computer worker is hired for a G job in the immoral, moral or neutral work condition, the computer worker would move a 2x3, a 2x2, or a 2x3 brick, respectively. That is, the computer worker would not reduce harm in the immoral work condition and would not increase the benefit in the moral work condition.

The market. Participants are randomly allocated to markets consisting of 6 workers and 4 "computer workers" who compete to be hired by 4 automated firms. Each worker provides two units of labor. We keep the wages for doing the N jobs fixed: for a first and a second N job in a market round, workers earn a wage of CHF 1.1 and CHF 0.50, respectively for doing the job.[26]

At the beginning of every market period, each worker decides whether or not to participate in the labor market for G jobs. In she decides to participate, she then (privately) provides two wage requests, one for each of the possible units of labor she can provide. Workers are not allowed to submit negative wage requests. The three computer workers always participate in the market for G job, and submit reservation wages of CHF 15 for each of the two jobs.

Firms are simulated by the computer. Each firm can hire up to one unit of labor per period. Labor demand is completely inelastic, which is known to the workers. The equilibrium prediction (assuming no concerns for morality) is that four different workers are hired to provide one G job each and the market wage is CHF 0.5.

As in Study 1, we use a uniform-price sealed-offer auction as the market mechanism. Once all six workers have submitted their wage requests for G jobs, the computer ranks them from lowest to highest. The computer also includes the wage requests of the computer worker. For every wage request that is ranked either 1st, 2nd, 3rd or 4th, the corresponding worker is hired to do a G job. The *market wage for G jobs* in a round is the 5th lowest wage request in that round, from among all the wage requests, including those submitted by computer workers. This mechanism clears the

[26] Instead of setting a fixed wage for neutral jobs, we could have created two separate markets with different wages: one for neutral jobs ($w(neutral\ job)$) and one for G jobs ($w(G\ job)$). However, we opted against having two markets for a couple of reasons. Firstly, having two markets would have made the experiment significantly more complex for participants. Secondly, the model that we use to guide our study (in Appendix C.2) suggests that what matters to workers is only the *relative* wage of G jobs, denoted as $\Delta w = w(G\ job) - w(neutral\ job)$. By fixing $w(neutral\ job)$ and allowing the market to determine $w(G\ job)$, we can measure the relative wage of G jobs, Δw, and identify a potential immorality work premium.

market in that, for the market wage, labor supply equals labor demand and all workers with wage requests below the market wage are hired. A worker's earnings in a round equal the market wage for G jobs times the units of G jobs provided by that worker (0, 1 or 2), plus the wage from doing neutral jobs. Workers who do not participate in the market receive CHF 1.6 for doing two neutral jobs.

The market repeats for a total of 12 periods. The composition as well as the type (immoral, moral or neutral) of each market is fixed across periods. At the end of each period, the computer reports the market wage and the total donations (or, "LEGO dots" in the neutral work condition) that have been replaced by the four workers hired to implement G jobs.[27] Moreover, as in Study 1, the report displays the picture of every worker in the market and summarizes information regarding each workers' outcomes across all periods. Specifically, subjects observe employment outcomes, wages and cumulative earnings for all workers in their market across periods, and can connect these to the other workers' identities through the photographs. After observing outcomes, the workers implement the two jobs for which they have been hired for.

Discussion. We create a market setting where, in each round, four G jobs will be implemented. If workers do not provide the G jobs, computers will do the jobs instead. Computer workers would not choose the "morally right" action at work (that is, reducing harm or increasing benefit, respectively). So, in the immoral work condition, workers cannot prevent the spread of misinformation or the transfer of donations from Everytown to the NRA. The only way to make a positive impact is by reporting low reservation wages for "immoral work," and when hired pay a cost of CHF 0.2 to decrease the relative donation to the NRA by CHF 0.4. Similarly, in ethical work, information about gun violence is always generated, and donations are transferred from the NRA to Everytown. Again, to have a positive impact, workers must report low reservation wages for "moral work," and when hired pay a cost of CHF 0.2 to decrease the relative donation to the NRA by CHF 0.4. From a purely consequential perspective, the immoral work and moral work conditions are similar. If workers are only motivated by the final donation to the NRA, we should not see any behavioral differences between the two treatment conditions. However, our model

[27] So, workers can't see the donations moved by each individual. To prevent disclosing this information through individual earnings, we display the earnings of every worker before subtracting potential costs based on their brick choice. One reason for not revealing individual donations is that although people often know the company and industry someone works for, workplace behavior is usually not publicly visible. Additionally, in our model, a worker's "social image" is tied to the company's image, *I(j)*, which is best represented by the total donation moved.

suggests that workers avoid personally doing immoral work, even when doing so is the only way of having positive impact.

D.3 Procedural details

All sessions took place at the UZH Laboratory for Experimental and Behavioral Economics at the University of Zurich in October through November of 2023. Participants were recruited from the joint subject pool of the University of Zurich and the ETH. Session consisted of 24 participants.[28] We conducted fifteen sessions, resulting in a total of 354 participants, allocated to 24 markets for immoral work, 23 markets for moral work and 12 markets for neutral work. The laboratory experiment was implemented with zTree (Fischbacher, 2007). Our study obtained ethical approval from the Human Subjects Committee of the Faculty of Economics, Business Administration and Information Technology at University of Zurich.

All instructions were delivered both on paper and with pre-recorded audio files. Instructions and materials are available at https://osf.io/4c6r7/. We preregistered the data collection and analysis. The preregistration plan can be found on https://osf.io/4c6r7/.

[28] In one session from the moral work treatment condition, we only had 18 participants (3 markets, each with 6 workers).

Appendix E – Construction of $\boldsymbol{\theta^{Sur}}$ and relationship between $\boldsymbol{\theta^{Sur}}$ and behavior in the laboratory and in the field

In this Appendix, we first give details about the online questionnaire that includes the psychological survey questions. We then construct an individual measure of concern for morality, θ^{Sur}, based on the answers to these survey questions. Finally, we show that this second measure of concern for morality correlates both with the comparable behavioral measure from the laboratory experiment (θ^{Exp}) and with outcomes in the laboratory labor market of study 1. This validation of θ^{Sur} is useful for future research, as it is based solely on survey questions that are easier to collect than the incentivized laboratory measure. Moreover, the comparison of θ^{Exp} and θ^{Sur} provides some evidence on the stability of moral concerns across time and contexts, which is necessary for heterogeneous moral concerns to persistently influence labor market behavior.

E.1 The online questionnaire

After asking asked subjects several questions about their future labor-market expectations (see section 7), subjects next encountered several multi-item scales intended to measure an individual's broad concern for morality and moral acts. These were:

1) **HEXACO-PI.** We administered 10 items from the short version of the HEXACO Personal Inventory (Ashton and Lee, 2009)[29] related to the factor "Honesty-Humility"—consisting of the four traits, sincerity (3 items), fairness (3 items), greed avoidance (2 items) and modesty (2 items). Every item describes a thought that a moral or immoral person might have and participants indicate the extent to which each thought reflects their own opinions.
2) **Protected Values.** The Protected Values scale (Gibson et al., 2013)[30] measures an individual's position regarding values that can be seen as inviolable, and not substitutable against money, and that are usually central to the person's identity. In our case, we adapted the Protected Values to a situation where a financial adviser can give bad investment advice to a client for personal benefit. First, 5 items assess the morality of this behavior (*Protected value 1*); second, 4 items examine how truthfulness matters in such a situation (*Protected value 2*).

[29] Ashton, M.C. and Lee, K. (2009) "The HEXACO-60: A Short Measure of the Major Dimensions of Personality," Journal of Personality Assessment, 91(4): 340-345.

[30] Gibson, R., Tanner, C. and Wagner, A. (2013) "Preferences for truthfulness: Heterogeneity among and within individuals," American Economic Review, 103(1): 532-548.

3) **Integrity and Work Ethics Test.** We used two items from an online test designed to allow firms to measure the integrity of job applicants (*Work ethics 1, Work ethics 2*). In each item, participants read fictitious dialogues between two characters with different opinions about a situation (e.g., calling in sick at work to enjoy a sunny day outside). Participants then rate with which character they have greater agreement.

4) **Charity attitude index.** We used a 9-item scale developed by Brashear et al. (2000)[31] in which participants rate statements regarding how important they perceive it is to help others in society and how positive and useful they perceive work done by charities.

In each case, subjects expressed agreement or disagreement with statements on either a 5-point or 7-point Likert scale. Descriptions and summary statistics of these survey scales are provided in Table E1.

Table E1: Description and summary statistics of survey scales

Variable	**Number of items**	**Mean (Sd)**	**Interpretation**
Protected value 1	5	0.75 (0.19)	1 = the person finds that the behavior of a banker who recommends sub-optimal assets to his clients because he has larger margins on them is: very outrageous, very blameworthy, very immoral, not at all acceptable and not at all praiseworthy.
Protected value 2	4	0.63 (0.18)	1 = the person thinks that truthfulness is a value that cannot be sacrificed.
Work ethics 1	1	0.38 (0.29)	1 = the person thinks that people are generally honest.
Work ethics 2	1	0.55 (0.36)	1 = the person thinks that calling sick to have a free day at work is really bad.
HEXACO sincerity	3	0.59 (0.19)	1 = the person is very sincere.
HEXACO fairness	3	0.68 (0.22)	1 = the person is very fair.
HEXACO greed avoidance	2	0.58 (0.23)	1 = the person is not at all greedy.
HEXACO modesty	2	0.66 (0.22)	1 = the person is very modest.
Charity attitude index	9	0.69 (0.13)	1 = the person's attitude towards charities is very positive.

Notes: Each subscale is constructed by taking averages over all items of the scale, and then normalized such that it lies between 0 and 1.

[31] Brashear, G., Green, L. and Webb, J. (2000) "Development and validation of scales to measure attitudes influencing monetary donations to charitable organizations," Journal of the Academy of Marketing Science, 28(2): 299-309.

We also asked subjects to provide unstructured responses stating beliefs about their future career trajectories—specifically, what work they expected to do after their studies and how much they expected to earn at the age of 40—and whether several non-profit organizations (including UNICEF) are worth supporting. We also collected additional personal characteristics using a short version of the Big Five (Gosling et al., 2003)[32] and several demographic characteristics.

Table E2: Items comprising $\boldsymbol{\theta^{Sur}}$ and their relationship to $\boldsymbol{\theta^{Exp}}$

	Factor loadings (weights for $\boldsymbol{\theta^{Sur}}$)	**Regression coefficient of $\boldsymbol{\theta}$**
	(1)	**(2)**
Protected value 1	0.664	0.540*** (3.66)
Protected value 2	0.708	0.352** (2.22)
Work ethics 1	0.213	-0.039 (-0.39)
Work ethics 2	0.252	0.042 (0.52)
HEXACO sincerity	0.482	0.362** (2.53)
HEXACO fairness	0.611	0.353*** (2.61)
HEXACO greed avoidance	0.477	0.225* (1.73)
HEXACO modesty	0.508	0.236* (1.79)
Charity attitude index	0.545	0.711*** (3.11)

*Notes: Each subscale is constructed by taking averages over all items of the scale, and then normalized such that it lies between 0 and 1. (1): Factor loadings from principal-component factor analysis of survey measures on θ^{Sur}. (2): Coefficient estimates of linear probability models. N = 237 for each regression (3 subjects did not complete the online-survey and are excluded). Dependent variable: being a high-theta type according to θ constructed from the behavioral task in the laboratory session. Independent variables: survey measures in [0,1], higher numbers indicate more morality. Robust standard errors; t-statistics in parentheses; * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.*

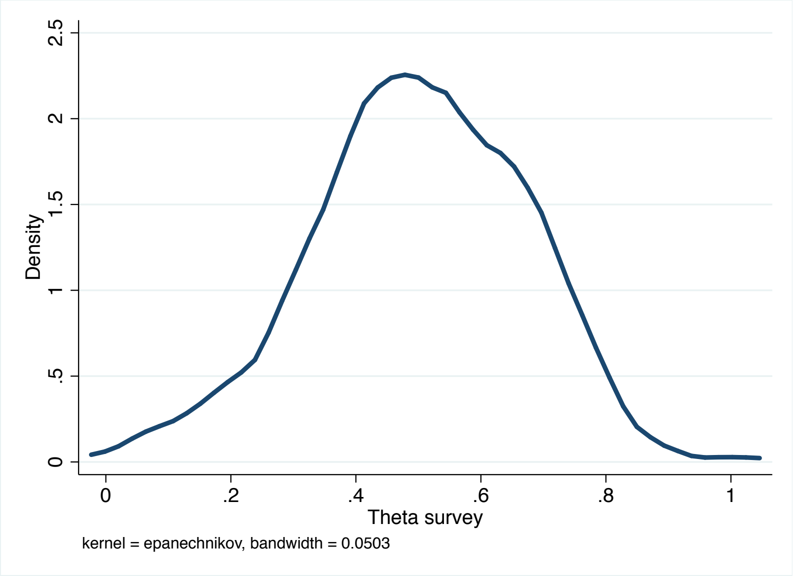
E.2 Constructing $\boldsymbol{\theta^{Sur}}$

We next discuss how we construct a survey-based measure of concern for morality, θ^{Sur}. Table E2 lists the 9 subscales from the morality measures. To reduce multiple measures of (presumably) related individual characteristics to a single dimension, we employ factor analysis. We perform a principal-component factor analysis and select the factor with the highest eigenvalue (eigenvalue

[32] Gosling, S.D., Rentfrow, P.J., and Swann Jr. W.B. (2003) "A very brief measure of the Big-Five personality domains," Journal of Research in Personality, 37: 504-528.

= 2.44) to represent θ^{Sur}.[33] Column 1 in Table E2 presents the corresponding factor loadings. We normalized θ^{Sur} such that it lies between 0 and 1; the resulting variable has a mean of 0.5 and a median of 0.5. Low values represent a low concern for morality. Figure E1 shows the distribution of θ^{Sur}. Given our interpretation of θ, we will often refer to subjects with a low θ^{Sur} as *immoral types* and subjects with high θ^{Sur} as *moral types*.

Figure E1: Probability distribution of θ^{Sur}



E.3 Does θ^{Sur} predict behavior in the laboratory?

To validate θ^{Sur}, we investigate how it correlates with behavior, roughly one week later, in the laboratory. Table E2, column 2, shows the coefficients from independent simple regressions of a subject's type measured by the behavioral laboratory task, θ, on each item comprising θ^{Sur}. The dependent variable is binary, indicating that a subject is a θ_H type according to θ. The results show a significant positive correlation between θ and all personality measures, except for *Work ethics 1* and *Work ethics 2*. Consistent with the positive relationship of the individual items, a regression of θ on θ^{Sur} shows a positive and significant relationship (coefficient=0.723, t=4.32, p<0.001); that is, a person who is characterized by a low concern for morality according to our survey-based measures is more likely to lie self-servingly in the behavioral measure in the experiment. The

[33] In Tables E3, E4 and E6, we show that our results are robust to different aggregation mechanisms. Specifically, we look at two alternatives: i) each of the nine survey measures is given equal weight and ii) the weight of the measures is determined by a regression of θ on the survey measures.

correlation between θ and θ^{Sur} is 0.270, and the Spearman's rank correlation is 0.253. We can also use a more continuous measure of θ, namely the number of lies and the expected payoff from lying; the correlations (Spearman's rank correlations) are 0.297 (0.271) for the number of lies and 0.282 (0.256) for the expected payoff from lying. These generally positive relationships suggest that θ and our survey-based measures capture a stable individual characteristic.

We next consider the extent to which θ^{Sur} also predicts participants' behavior in the laboratory labor market, particularly in the immoral work condition.[34] Results from a linear regression of the employment rate on θ^{Sur} indicate that those participants with the lowest concerns for morality (that is, participants with $\theta^{Sur} = 0$) are 43.9 percentage points more likely to be hired in markets for immoral work than participants with the highest possible concern for morality (that is, with $\theta^{Sur} = 1$). This difference is marginally statistically significant (p=0.057, see Table E3, column 1). A less noisy measure of subjects' market behavior is their actual choices. Results from a hurdle model indicate that those subjects with the lowest concerns for morality are 52.1 percentage points more likely to participate in markets for immoral work by submitting a wage request than individuals with the highest possible concerns (p=0.015, see Table E4). In the neutral work condition, as expected, we find no substantial differences in employment rates or in labor market behavior between the two types.

Table E3: Relationship between $\boldsymbol{\theta^{Sur}}$ and outcomes in the experimental labor market

Dependent variable:	**Employment rate**		
	(1)	(3)	(5)
Type survey ($\boldsymbol{\theta^{Sur}}$)	-0.439* (-1.97)	-0.546* (-1.74)	-0.638* (-1.87)
Neutral work (N)	0.046 (0.37)	-0.104 (-0.45)	-0.133 (-0.59)
$\boldsymbol{\theta^{Sur}}$ * N	0.431 (1.65)	0.558 (1.54)	0.648* (1.70)
Aggregation $\boldsymbol{\theta^{Sur}}$	Factor Analysis	Equal weight	Theta-Exp
N	3,555	3,555	3,555
R²	0.080	0.077	0.078
p-value: $\boldsymbol{\theta^{Sur}}$+ $\boldsymbol{\theta^{Sur}}$*N = 0	0.953	0.945	0.953

*Notes: Coefficient estimates of linear regression models. Models differ in how we construct θ^{Sur} from the nine psychological survey measures. Column (1) reports our main results, using factor analysis to aggregate the psychological measures. Column (2) gives the result if equal weight is given to each measure instead. Column (3) gives the results if weights are determined by a regression of the survey measures on θ. Other independent variables: Immoral work is in {0, 1}, θ^{Sur} is in [0,1], where higher numbers indicate more morality. Standard errors clustered at market level; t-statistics in parentheses; * - p < 0.1; ** - p < 0.05; *** - p < 0.01.*

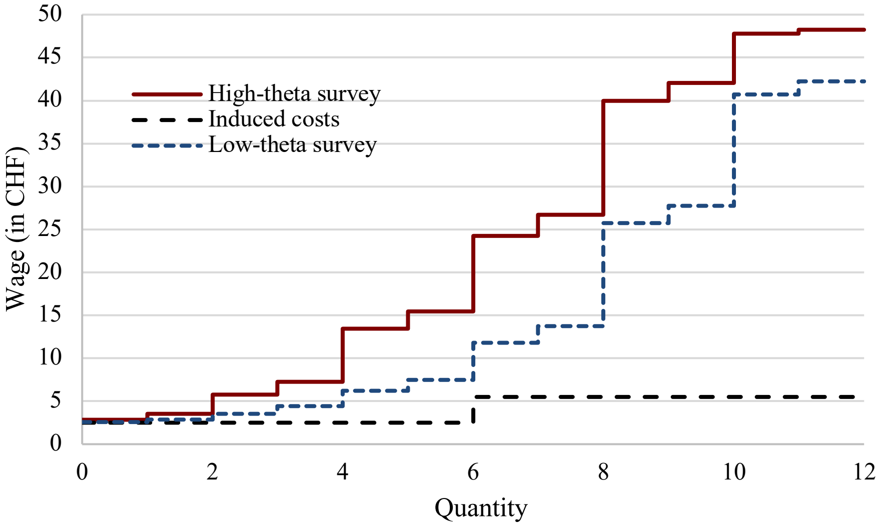
[34] Figure E2 displays the labor supply in a (simulated) labor market with only low θ^{Sur} or only high θ^{Sur}types.

Table E4: Relationship between participation decision/reservation wage and θ^{Sur} (Hurdle model) in the immoral work condition

Dependent variable:	Participate	Reservation wage	Participate	Reservation wage	Participate	Reservation wage
	(1)	(2)	(3)	(4)	(5)	(6)
Type survey (θ^{Sur})	-1.614**	0.432	-1.971**	0.571	-2.177**	0.373
	(-2.44)	(0.60)	(-2.08)	(0.61)	(-2.13)	(0.33)
Constant	1.326***	3.674***	1.806***	3.515***	1.841***	3.657***
	(3.98)	(10.39)	(2.99)	(6.12)	(3.01)	(5.25)
Sigma		2.679***		2.679***		2.680***
		(7.71)		(7.72)		(7.70)
Aggregation θ^{Sur}	Factor Analysis	Factor Analysis	Equal weight	Equal weight	Theta-Exp	Theta-Exp
N	2’475	1’711	2’475	1’711	2’475	1’711
LL (pseudo)	-1478.3	-4114.1	-1488.2	-4114.2	-1490.9	-4114.6

*Notes: Estimates from Craggs double-hurdle model: Regressions (1), (3) and (5) are probit models, regressions (2), (4) and (6) are truncated linear regressions (truncated from above at 50 CHF). Regressions differ in how θ^{Sur} is constructed from the nine psychological survey measures. Columns (1) and (2) report our main results, using factor analysis to aggregate the psychological measures. Columns (3) and (4) give the result if equal weight is given to each measure instead. Columns (5) and (6) give the results if weights are determined by a regression of the survey measures on θ. θ^{Sur} is in [0,1], where higher numbers indicate more morality. Sample incudes only subjects from the immoral work condition. (In the neutral work condition, the coefficient of θ^{Sur} is not significant for any of the above specifications.) Standard errors clustered at market level; z-statistics in parentheses; * - $p < 0.1$; ** - $p < 0.05$; *** - $p < 0.01$.*

Figure E2: Labor supply for immoral work in the laboratory for different types (θ^{Sur})



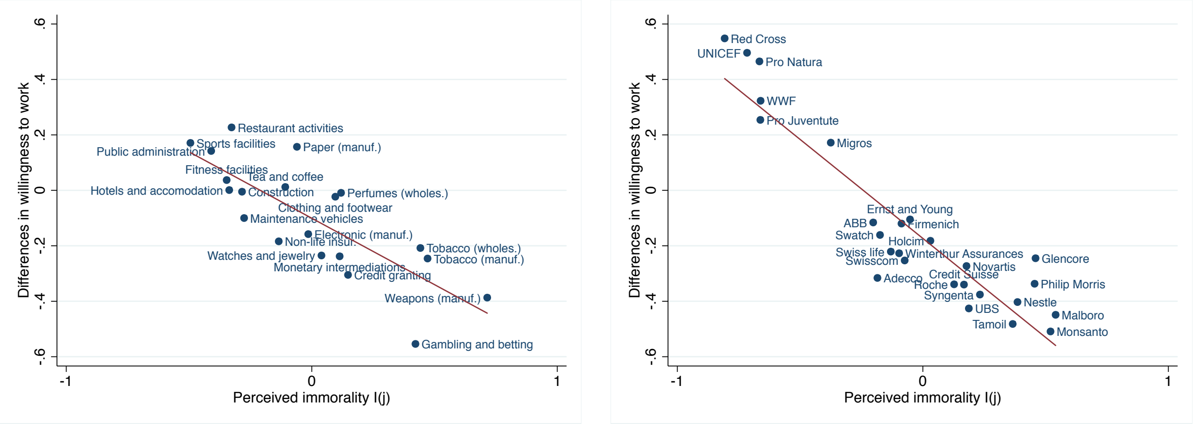
Notes: High-theta survey is 1 if θ^{Sur} is lower than the median of θ^{Sur}. Labor supplies are calculated with simulations: 6 labor market decisions (first and second wage request) of high-theta survey (or, low-theta survey) types are randomly drawn (without replacement) from our sample. We then calculate the labor supply for this group of people. We repeat this 1000 times (with replacement) and take the average of these 1000 individual labor supplies.

E.4 Does θ^{Sur} predict stated real-world labor market preferences?

In the following, we replicate the analysis in Section 7.2 using θ^{Sur} instead of θ. That is, we investigate whether the perceptions of industry and firm immorality interact with our subjects' measured concern for morality θ^{Sur}. As in Section 7.2, we normalized subjects' stated willingness to work for firms and industries, such that they take values between 0 (*not at all willing*) and 1 (*very much willing*).

The vertical axis of Figure E3a plots the difference in willingness to work in an industry between subjects classified as moral or immoral, according to θ^{Sur}. The strong negative relationship across both figures indicates that subjects classified as immoral using our survey-based (θ^{Sur}) are, on average, more willing to work for industries that others perceive as immoral.

Figure E3: Correlation between the difference in willingness to work between moral and immoral types and perceived immorality of industries/firms



(a) Industries, θ^{Sur}

(b) Firms, θ^{Sur}

Source: Survey study (Perceived immorality), online survey (Willingness to work, θ^{Sur})
Notes: Differences in willingness to work: Coefficient estimates of linear regression models of the participants' willingness to work for different industries (a) or firms (b) on θ^{Sur}. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: θ^{Sur} in [0,1]. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral.

Table E5, columns (1) and (2) provide statistical evidence of the relationships in Figure E3a. The dependent variable is a subject's willingness to work for an industry, while the explanatory variables include the perceived industry immorality ($I(j)$, obtained from a separate group of respondents), the subject's concern for acting morally (θ^{Sur}) and the interaction of these two terms. While there is little evidence of a systematic difference in willingness to work for neutral industries between moral and immoral types, subjects' moral types have much stronger predictive power for their willingness to work in industries perceived as immoral. This pattern is significant at the 1%-level, and is robust to controlling for subjects' gender, age, Swiss nationality, area of study, mean industry wages, industry size (number of employees), and industry sales.

Table E5: Regressions of willingness to work for diverse industries and firms on perceived immorality and moral types

Dependent variable:	**Willingness to work for industry *j***		**Willingness to work for firm *j***	
	(1)	(2)	(3)	(4)
Perceived immorality (I(j))	-0.050 (-0.77)	-0.043 (-0.74)	0.114 (1.56)	0.110 (1.50)
Type from survey ($\boldsymbol{\theta^{Sur}}$)	-0.101* (-1.71)	-0.107* (-1.72)	-0.173** (-2.28)	-0.211*** (-2.85)
$\boldsymbol{\theta^{Sur}}$ * I(j)	-0.479*** (-5.22)	-0.479*** (-5.24)	-0.731*** (-8.80)	-0.722*** (-8.53)
N	4715	4715	5064	5064
Control variables	No	Yes	No	Yes

*Notes: Coefficient estimates of linear regression models. Dependent variable: Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Observations where subjects did not know the firm ("I don't know this organization") or did not fill out the questionnaire are excluded. Independent variables: θ^{Sur} (in [0,1]), Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study, average wage industry 2016 (SLFS; only for industries), industry size 2016 (STATENT; only for industries), industry sales 2015 (Value Added Tax Statistics; only for industries). Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*

We repeat the same analysis using data on subjects' willingness to work for our selection of well-known *firms* in Switzerland. The vertical axis of Figure E3b plots the difference in willingness to work for firms between subjects classified as moral and immoral according θ^{Sur}. Again, subjects classified as immoral are, on average, more willing to work for firms perceived as immoral. Table E5 confirms this relationship in columns (3) and (4): subjects less concerned with moral behavior are more willing to work for firms that other people rate as more immoral (p<0.01). Hence, we find that firms perceived as immoral more attractive to potential workers with a lower concern for morality. As we show in Table E6, the differential willingness to work for immoral firms and industries by moral and immoral types does not depend on how we construct θ^{Sur}.

Table E6: Regressions of willingness to work for diverse industries and firms on perceived immorality and moral types, robustness checks aggregation $\boldsymbol{\theta}^{\mathbf{Sur}}$ (Study 1)

Dependent variable:	**Willingness to work for industry *j***				**Willingness to work for firm *j***			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Perceived immorality (I(j))	0.089 (1.00)	0.096 (1.17)	0.122 (1.33)	0.129 (1.43)	0.396*** (3.65)	0.391*** (3.58)	0.337*** (4.16)	0.327*** (4.00)
Type survey ($\boldsymbol{\theta}^{Sur}$)	-0.154* (-1.93)	-0.149* (-1.77)	-0.131 (-1.49)	-0.158* (-1.68)	-0.200* (-1.84)	-0.267*** (-2.58)	-0.243** (-2.21)	-0.298*** (-2.69)
$\boldsymbol{\theta}^{Sur} * \mathbf{I(j)}$	-0.583*** (-5.05)	-0.583*** (-5.06)	-0.671*** (-4.95)	-0.671*** (-4.98)	-0.998*** (-7.83)	-0.990*** (-7.60)	-0.961*** (-8.54)	-0.944*** (-8.26)
Aggregation $\boldsymbol{\theta}^{Sur}$	Equal weight	Equal weight	Theta-Exp	Theta-Exp	Equal weight	Equal weight	Theta-Exp	Theta-Exp
N	4’715	4’715	4’715	4’715	5’064	5’064	5’064	5’064
Control variables	No	Yes	No	Yes	No	Yes	No	Yes

*Notes: Coefficient estimates of linear regression models. Observations where subjects did not know the firm (“I don’t know this organization”) or did not fill out the questionnaire are excluded. Independent variables: Models differ in how we construct θ^{Sur} from the nine psychological survey measures. Columns (1), (2), (5) and (6) give the result if equal weight is given to each measure. Columns (3), (4), (7) and (8) give the results if weights are determined by a regression of the survey measures on θ_L^{Exp}. θ^{Sur} is in [0,1] where higher numbers indicate more morality. Willingness to work is in {0, 0.25, 0.5, 0.75, 1} where 0 means not at all willing to work, 0.5 means indifferent and 1 means really much willing to work. Perceived immorality is in [-1, 1] where -1 means very moral, 0 means neutral and 1 means very immoral. Control variables: age, gender, Swiss nationality, subject of study, average wage industry 2016 (SLFS; only for industries), industry size 2016 (STATENT; only for industries), industry sales 2015 (Value Added Tax Statistics; only for industries). Standard errors clustered at individual and industry/firm level (Cameron, Gelbach and Miller, 2011); z-statistics in parentheses; * p < 0.1; ** p < 0.05; *** p < 0.01.*