

Ruffle, Bradley; Tobol, Yossi

Working Paper
Screening for Honesty

IZA Discussion Papers, No. 8286

Provided in Cooperation with:
IZA – Institute of Labor Economics

Suggested Citation: Ruffle, Bradley; Tobol, Yossi (2014) : Screening for Honesty, IZA Discussion Papers, No. 8286, Institute for the Study of Labor (IZA), Bonn

This Version is available at:
<https://hdl.handle.net/10419/101902>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

IZA DP No. 8286

Screening for Honesty

Bradley J. Ruffle
Yossef Tobol

June 2014

Screening for Honesty

Bradley J. Ruffle

*Wilfrid Laurier University and
Ben-Gurion University*

Yossef Tobol

*Jerusalem College of Technology,
Carmel Academic Center and IZA*

Discussion Paper No. 8286
June 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Screening for Honesty¹

We report the results of a field experiment on honesty conducted on 427 Israeli soldiers fulfilling their mandatory military service. Each soldier rolled a six-sided die in private and reported the outcome to the unit's cadet coordinator. For every point reported, the soldier received an additional half hour early release from the army base on Thursday afternoon. We find that the higher a soldier's military entrance score, the more honest he is on average. Moreover, to the extent that honesty is a valued trait, regression discontinuity analysis reveals that the Israeli military has optimally set the threshold score to qualify to be an officer. Our results bear important implications for the design of screening tests that evaluate employee honesty.

JEL Classification: C93, M51

Keywords: experimental methods, honesty, personnel selection, soldiers, high non-monetary stakes, regression discontinuity design

Corresponding author:

Yossef Tobol
School of Management
Jerusalem College of Technology
Jerusalem 91160
Israel
E-mail: toboly@gmail.com

¹ We thank Yuval Arbel, Ofer Azar, Ronen Bar-El, Leif Danziger, Naomi Feldman, Shachar Kariv, Jonathan Mamujee, Mattia Pavoni, Chet Robie, Tata Pyatigorsky-Ruffle, Jonathan Schulz, Ze'ev Shtudiner, Michal Kolodner-Tobol, Ro'i Zultan and especially Danny Cohen-Zada for helpful comments. We also are grateful to Capt. Sivan Levi and Meytal Sasson for research assistance, Capt. Itamar Cohen for facilitating these experiments and all of the commanding officers for granting us access to their units.

1. Introduction

Theft and fraud reduce annual company revenues by 5% on average, amounting to \$3.7 trillion in losses worldwide (Association of Certified Fraud Examiners 2014, p. 4). The U.S. retail industry alone loses \$53.6 billion a year to employee theft. Moreover, employee theft is on the rise due to poor pre-employment screening and a decline in supervision (Brooks 2013). Can companies screen better for honest employees? Polygraph tests have been shown to be unreliable and their use by employers is unlawful in North America and Europe. Neuroscience-based lie-detection technologies remain unproven and in their infancy. Consequently, many employers continue to rely on written personality tests consisting of self-report questions and job interviews led by the company's human resource personnel to evaluate job candidates' honesty. However, a considerable body of research casts doubt on the usefulness of these methods (see, e.g., Morgeson et al. 2007 for a survey and Ones et al. 2007 for a rejoinder).

In this paper, we test the effectiveness of a method employed by the Israeli Defense Forces (IDF) that combines a personal interview conducted by highly select professionals with extensive specialized training in reading body language and evaluating honesty and integrity along with a psycho-technical written exam to assess, among other traits, each candidate soldier's honesty. To this end, we conduct a field experiment on a representative sample of 427 soldiers and show that the IDF's measure predicts the level of honesty observed in our experiment. What is more, regression discontinuity analysis reveals that the IDF has set optimally its cutoff score to qualify to be an officer. Based on these findings, we suggest that employers that value honesty reexamine their current practices and consider adopting some of the key features in the IDF's screening tests.

It would be difficult to exaggerate the importance of honesty in a relationship. Trust is predicated on honesty. An honest relationship encourages those parties involved to invest in the relationship, whereas an absence of honesty dooms the relationship. In the workplace, honesty is essential for group-oriented tasks, especially since they are usually not well-defined or contractible. Concerns of dishonesty also beset principal-agent problems. When one principal is responsible for numerous agents and is unable to monitor their actions, concerns surrounding the agents' honesty become even more acute. In difficult economic times, employers typically invest fewer resources in monitoring

technologies and, with wages being frozen or sometimes even cut, employees are lured into stealing from their employers. Absentee company management cannot monitor its employees' honesty and integrity.

Military units are also susceptible to dishonesty among its soldiers. The hierarchical organizational structure inherent in a nation's military inevitably means that a commanding officer who assigns a duty or issues an order to a soldier has no opportunity to verify whether the soldier has completed the assigned task. Yet a well-functioning military relies on honesty between its troops and even seeks this trait when recruiting and promoting soldiers.

Indeed, honesty is among the highest declared values for Israeli soldiers and part of the creed of the Israel Defense Forces (IDF). Its compulsory entrance exam screens for honesty through a one-on-one interview with a highly skilled professional trained to evaluate interviewees' honesty as well as through numerous questions and checks for response consistency on a written psycho-technical test. We use the methods of a laboratory experiment on a population in the field (in the terminology of Harrison and List (2004), ours is an artefactual field experiment) to assess the IDF's effectiveness at categorizing incoming soldiers on the basis of their honesty. We extend Fischbacher and Föllmi-Heusi's (2013) innovative die-rolling paradigm to the field where each soldier rolls a 6-sided die in private and reports the outcome to his army unit's cadet coordinator. The payment is an additional half-hour early release from the army base on Thursday for each point reported on the die. We test whether a number of the soldier's characteristics, including his military entrance score which is partially determined by his assessed degree of honesty, predict his honesty in our experiment.

Our experiment differs from a growing experimental literature on honesty in two important respects. One distinctive feature is the novelty of our subject pool and the payment. Unlike student subject pools or even most field experiments targeted at a particular population, soldiers completing their mandatory military service constitute a representative cross-section of society as a whole.² More interestingly and to the best of

² Abeler et al. (2014) conduct an experiment on honesty by telephone on a representative sample of German respondents. Other experiments conducted on soldiers are: Goette et al. (2012) who compare the in-group cooperativeness and willingness to punish of extant groups of Swiss soldiers with those of randomly formed groups of soldiers. Lahav et al. (2011) distribute questionnaires on trains traveling between major

our knowledge, ours is the first experiment to examine honesty toward one's employer. Soldier subjects in our experiment cheat their boss with whom they interact on a daily basis, rather than an anonymous firm (e.g., Levitt 2006; Pruckner and Sausgruber 2013), anonymous subjects (e.g., Gneezy, 2005), wait staff at a restaurant (Azar et al. 2013), or the experimenter (e.g., Fischbacher and Föllmi-Heusi 2013).³ This novel feature and, as our findings show, the apparent success of the IDF at sorting soldiers according to their degree of honesty have practical implications for firms that aim to hire honest employees.

For over a century, personality tests have been employed to aid in hiring decisions, job placement and worker evaluation. With the advent of the personal computer to administer these tests, their usage has exploded over the past few decades. These tests consist of numerous self-report questions designed to evaluate a spate of personality traits, including reliability, creativity, motivation, sociability and honesty. Their prevalence has given rise to industries devoted to designing them, computer applications to administer and grade them, and books and websites providing guidance to test-takers. The hidden agenda of many of these questions is to evaluate the job candidate's honesty. For example, genuine admission of personal faults and past errors are expected in response to “what don't people like about you?” and “what mistakes have you made?” A question ostensibly about worker motivation, “do you believe that work is by far the most important thing in your life?” is another insidious attempt to assess the respondent's honesty. The most common test question – also about honesty – is even trickier: respondents are asked to indicate on a scale of 1 (strongly agree) to 5 (strongly disagree) the extent to which they agree with the statement: “I have never told a lie”.

Ghiselli and Barthol (1953) examined 113 studies published from 1924 to 1950 dealing with the validity of personality measures in employee selection and, paradoxically, found them to be effective predictors for jobs where such personality

Israeli cities to soldiers, teenagers and university students and show that soldiers have higher subjective discount rates than non-soldiers. In a companion paper, Ruffle and Tobol (2014) show that temporally distancing decisions from the receipt of payment increases honest reporting. Specifically, soldiers who participated in the die-rolling experiment on earlier days of the week reported low outcomes on average than those who participated closer to the end of the week. Soldiers' military entrance scores served merely as a control variable in the analysis and was not explored in any depth. In the current paper, we focus on the relationship between military entrance scores and honesty and, to the extent that the Israeli military prizes honesty, whether it has set optimally the entrance score cutoff required to become an officer.

³ To a lesser extent they also cheat their colleagues because a soldier who leaves the army base early necessitates that his uncompleted duties are distributed among those soldiers who remain behind.

factors might be considered to be of limited importance (e.g., clerks) and ineffective for jobs where personality would be expected to be vital (e.g., supervisors). Guion and Gottier (1965) conducted a meta-analysis summarizing 12 years of research and conclude that “there is no generalizable evidence that personality measures can be recommended as good or practical tools for employee selection” (p. 159). More recently, two meta-analyses (Barrick and Mount 1991; Tett et al. 1991) endorse the validity of personality tests for personnel selection. Their findings spawned a renewed interest in research on personality testing in the past two decades. Morgeson et al. (2007) review over 7000 manuscripts on the usefulness of personality tests and conclude that they have low predictability of job performance and that alternatives to these self-report measures should be sought. Our findings suggest that the IDF’s combined written test and particularly the interview component provide one such compelling alternative.

2. Experiment Design, Procedures and Sample

Our sample consists of 427 soldiers from 27 different permanent and provisional military bases throughout Israel and 15 distinct army units.⁴ All soldiers were serving in their first of three years of required military service. Within each participating army company, all soldiers took part in our experiment. The experiments were conducted between December 28, 2010 and June 19, 2011, a period of relative quiet in Israel as it was not involved in any wars or military confrontations.

All of the experiments were conducted just prior to the soldiers' breakfast hour in the dining hall. The cadet coordinator (CC) of the participating army unit called each soldier by name one-at-a-time to a room or large tent with two entrances/exits located on the army base and used for the purpose of the experiment. Each participating soldier entered through one designated entrance. The CC then read from a script the rules of the experiment to the soldier. Namely, the soldier was told that s/he would be asked to roll a 6-sided die in private and then to report the outcome to the CC. For each point on the die, the soldier would be released on Thursday half an hour ahead of the scheduled time. To

⁴ With the exception of our purposeful oversampling of religious companies, we view our sample of soldiers as representative of the overall population of Israeli soldiers. In fact, in section 3.2 we will see that the distribution of military entrance scores of soldiers in our sample mirrors the overall distribution.

avoid any possible confusion, the exact payment in the form of hours of early release for each of the six possible outcomes was enumerated. The soldier was told that, after all soldiers in the unit had completed the experiment, the CC would submit the list of early release times to the unit commander who had approved the experiment and the terms of early release.

The soldier was handed a 6-sided die and proceeded to a table at the other side of the room or tent, out of sight of the CC. After rolling the die in private, the soldier returned to the CC to report the outcome. Finally, the soldier completed a brief post-experiment questionnaire (included in the Appendix), submitted it to the CC and was directed to proceed to the dining hall through the door or tent opening designated as the exit and through which he had not entered. The distinction between the two doors or tent openings as entrance and exit was maintained to prevent soldiers from having contact with others who had not yet participated in the experiment. The CC called in the next soldier according to the list and so on until all soldiers in the unit had completed the experiment.

The entire experiment including the questionnaire took about seven minutes for each soldier. In view of the value soldiers attribute to an early release of half an hour (median = 30 NIS, see rows 2 and 3 of Table 1) and of three hours (median = 100 NIS, row 4 of Table 1), the experimental payment can be deemed salient.⁵

<insert Table 1 about here>

3. Results

3.1 Overall Distribution

The distribution of reported die outcomes for our entire sample of soldiers (N=427) is displayed in Figure 1. If all soldiers reported the truth, we would expect a uniform distribution of reported die outcomes. This hypothesis is soundly rejected (Pearson chi-

⁵ Consider the following back-of-the-envelope calculation. The average subject reported a die outcome of 3.87 (see row 1 of Table 1), equivalent to 1.97 hours early release. If we assume, for simplicity, that the median willingness to pay increases linearly with each additional half hour of early release, then the median willingness to pay for 1.97 hours equals 70.2 NIS for seven minutes of work. Contrast this with combat soldiers' monthly wage of 700 NIS a month and non-combat soldiers' monthly salary of between 300 and 500 NIS, depending on their job. At the time of the experiments, 3.5 NIS equaled \$1 USD.

square test $\chi^2(5) = 16.2, p = .001$).⁶ Soldiers clearly inflate their reported outcomes, but do not profit maximize.⁷ What is more striking is the observed decline in frequency from 5 to 6. While Fischbacher and Föllmi-Heusi (2013) also report incomplete cheating, they still register a higher percentage of subjects who report the highest outcome, whereas we witness a sharp decline. One explanation for our observed decline in reported 6s is that payments are publicly observable. A soldier seen leaving the base three hours early on Thursday may be concerned that his peers will view him as dishonest.

<insert Figure 1 about here>

3.2 Honesty and Military Entrance Score

Months prior to recruitment to the Israeli military, every candidate soldier is evaluated on the basis of his or her educational background and a series of computerized psycho-technical exams. Numerous questions on the psycho-technical exams are designed to evaluate the soldier's honesty by, for instance, framing the same question in multiple ways or otherwise asking as many ten variations of the same question to test for consistent responses. In addition, males undergo a lengthy personal interview in which the candidate's honesty is evaluated through several channels. First, female interviewers aim to assess the male soldier's "body language, to identify lies and individuals who are unreliable" (Hebrew Wikipedia under "recruitment to the Israeli military").⁸ What is more, candidate soldiers are asked a battery of questions, the answers to which are either already known (e.g., "Have you been caught stealing or otherwise been in trouble with the police?"), can be verified by contacting the appropriate authority (e.g., "Have you ever skipped school?", "What would your high school teachers say about you?") or can

⁶ Further evidence against of uniformly distributed die outcomes comes from the frequencies of reported 1s and 2s, both significantly less than the percentage of 16.67% expected from a uniform distribution ($p < .001$ from one-sided Binomial tests in both cases). At the same time, the frequencies of 4s and 5s, are significantly greater than 16.67% ($p = .04$ and $p < .001$, respectively). Only the reported frequencies of 3s and 6s cannot be rejected as significantly different from 16.67% ($p = .13$ and $p = .38$, respectively).

⁷ Incomplete cheating appears to be a robust finding in the emerging literature on cheating regardless of whether the die-rolling paradigm (e.g., Shalvi et al. 2011, Fischbacher and Föllmi-Heusi 2013, Hao and Houser 2013) or some other experimental method is used (e.g., Gneezy 2005, Charness and Dufwenberg 2006, Erat and Gneezy 2012).

⁸ Interestingly, Daniel Kahneman developed in large part the structured interview protocol, which remains largely intact to this day (Kahneman 2002).

be cross-checked with the candidate's responses on the written exam (e.g., "Have you ever used drugs?", "Are you seeking the ideal partner or willing to compromise?").

For male soldiers, the final test score (known as *kaba* in Hebrew and to be subsequently referred to as such for brevity) is made up of the interview (33%), the psycho-technical exams (50%) and the candidate's educational background (e.g., school attended, absences, any recorded discipline issues) and achievements (e.g., grades, clubs, distinctions) (17%). Women do not undergo the interview. Instead, their *kaba* is based on the psycho-technical exams (60%) and their educational background and achievements (40%).

An individual's *kaba* determines the unit and job to which he is assigned for his military service. The *kaba* scores range from 41 to 56, with special significance given to 52, the required cutoff to qualify for an officer's course. It is noteworthy that a *kaba* of 54 constitutes the minimum score to qualify to be an interviewer, with most having attained a rare perfect score of 56. In fact, the criteria for selecting interviewers are considered to be more stringent than any other unit in the IDF (Lerer 2009, pp. 20-32).

Through each unit's commanding officer and unbeknown to the soldier, we obtained every participating soldier's *kaba*. Figure 2 displays the distribution. The most unusual feature of the distribution is the paucity of near-miss scores of 50 and 51. At the same time, the highest frequency is associated with 52 (the threshold to qualify for an officer). These features suggest that the Israeli military wishes to enlarge the pool of eligible officers. To do so, it uses the subjective components of the *kaba* score, namely, the personal interview and the evaluation of the candidate's educational background, to boost to 52 the scores of soldiers deemed suitable.

The distribution of entrance scores reveals that 48.2% of soldiers in our sample qualify to be officers. Interestingly, Lerer (2009, p. 165) reports an identical figure (48%) for the fraction of soldiers with a *kaba* of 52 or higher in 1995 (the most recent year for which he obtained data).⁹

<insert Figure 2 about here>

The weighted scatter plot in Figure 3 displays the distribution of reported die outcomes for each *kaba* score. The size of the circle reflects the number of observations

⁹ The Israel Defense Forces do not make publicly available the distribution of military entrance scores.

at this *kaba*-die outcome pair. The figure also includes two regression curves fitted to the data. The OLS line displays a negative slope, in other words, a positive relationship between a soldier's *kaba* and his honesty. So as not to impose a linear fit on the data, we also estimate a kernel regression (shown in red in Figure 3). This curve too is negatively sloped up to and including a *kaba* of 54, after which a modest upturn is observed.¹⁰

The OLS regression in (1) of Table 2 pinpoints the magnitude and statistical significance of the negative relationship: for every additional point a soldier obtained on his *kaba*, he reports 0.13 points less on the die outcome ($p < .01$).¹¹ Regression (2) demonstrates that this highly significant negative relationship between a soldier's reported die outcome and his *kaba* remains robust to the inclusion of numerous controls. These controls include the soldier's self-reported willingness to pay for one half-hour early release from the military base, indicator variables for whether the soldier is female, religious (and an interaction term between the two) and from a city, and the soldier's response to a question about the extent to which he generally tells the truth.¹² The majority of the controls do not differ significantly from zero in this or any subsequent regression. The two exceptions are female soldiers who report weakly significantly higher die outcomes than males in this regression only.¹³ The mean die outcomes reported by males and females do not differ significantly from one another in any of the other regressions. Soldiers from cities (defined by the Israeli Central Bureau of Statistics as settlements with more than 20,000 residents) report significantly higher die outcomes than those from rural areas in this and all subsequent regressions.

¹⁰ Warner and Pleeter (2001) also observe unique behavior among the two highest categories of entrance exam scores in the U.S. military. In particular, they exploit a natural experiment conducted by the U.S. Department of Defense to reduce military personnel in which mid-career personnel were offered the choice between a lump-sum separation payment and an annuity valued at considerably more in present terms. Personnel belonging to the top groups display lower rates of discount (i.e., more patience) than their peers, as evidenced by their higher likelihood of preferring an annuity to a lump-sum retirement payment.

¹¹ The regressor is expressed as soldier *i*'s *kaba* minus 52 for ease of interpretation. Thus, the constant of 3.76 refers to the average die outcome reported by a soldier with a *kaba* of 52.

¹² Also included but not shown are measures of the soldier's military unit and military base peer effects (neither of these measures is significantly different from zero in any of the regressions), as well as indicators for the day of the week on which the experiment was conducted with Sunday as the omitted day (all of the other days of the week are positive and significantly different from zero in all regressions and are discussed in detail, along with the peer effects variables, in Ruffle and Tobol 2014).

¹³ Contrast this with Dreber and Johannesson (2008) who find that men are more likely than women to send deceptive, self-serving messages to their partner in a sender-receiver game modeled after Gneezy (2005).

Surprisingly, a soldier's willingness to pay for early release is unrelated to a soldier's decision to claim additional early release ($p=.12$). According to Table 1, the mean willingness to pay for a half-hour early release is 42.7 NIS. A standard deviation of 67.2 that exceeds the mean by more than 50% and a median of 30 NIS both attest to outliers. If we exclude observations that deviate from the mean by more than two standard deviations, the median remains unchanged, while the mean and standard deviation drop to 33.9 NIS and 28.1 NIS respectively ($N=412$). Regression (3) excludes these 15 outlying observations, but is otherwise identical to (2). The coefficient on the willingness-to-pay variable increases sevenfold to .007 and is highly significant ($p=.01$). Still it remains economically unimportant: a one-standard deviation in a soldier's willingness to pay increases his reported die outcome by only 0.20 points on average. At the same time, the *kaba* coefficient of -0.11 in (3) remains highly significant.

<insert Figure 3 about here>

<insert Table 2 about here>

There are two possible explanations for the positive association between a soldier's honesty and his entrance score. First, with its emphasis on the value of honesty among its soldiers, the Israeli military successfully discerns soldiers' degree of honesty and adjusts soldier's entrance scores accordingly. Second, soldiers behave reciprocally toward the military: soldiers know their entrance scores and use our experiment to reward or to punish the military for a high or low score received, respectively.

To distinguish between these two explanations, we introduced two questions midway through the data collection process: 217 soldiers were asked whether they knew their entrance score and, if so, what it was (see questions 8 and 9 of the Appendix). In our sample, 15% of soldiers admitted to not knowing their entrance score. Among those who claimed to know their score, only 47% indicated the correct score. Of the 53% of the soldiers who incorrectly guessed their score, the vast majority 92/97 overestimated it. The average guess among soldiers who mistakenly guessed their score was 1.98 points higher than the actual score ($s.d.=.138$).¹⁴

¹⁴ Our observations that as many as 40% ($.47*(1-.15)$) of soldiers correctly guessed their *kaba* and those who guessed incorrectly were off by "only" 2 points on average are not surprising. Before entering the military, every recruit provides a preference ordering over military units in which he wishes to serve. Since

If soldiers use our experiment to express their appreciation for a high entrance score and the enhanced opportunities that such a score furnishes or to express their dissatisfaction with their low score, then we would expect a soldier's guess about his score to predict significantly his reported die outcome. Regression (4) replaces a soldier's true entrance score with his guess. The estimated coefficient is actually positive (0.09), but not significantly different from zero ($p=.24$). The coefficient remains close to and not significantly different from zero when our set of control variables is included in (5). By contrast, the parallel regressions in (6) and (7) that use the same sample of 185 soldiers, replacing their guess with their true score, reveal highly significant coefficients of -0.18 ($p=.001$) and -0.13 ($p=.05$), respectively. In short, reciprocity cannot explain the observed relationship between honesty and soldiers' test scores.

3.3 Regression Discontinuity Design

Just because a soldier's actual military entrance score predicts his degree of honesty does not imply that the threshold score of 52 to qualify to be an officer has been optimally set. To explore the optimality of the threshold, we employ a regression discontinuity (RD) design (see Lee and Lemieux 2010 for an introduction). Eligibility to be an officer in the Israeli military depends deterministically and discontinuously on one's military entrance score. A score of 52 or more qualifies one to be an officer, whereas less than 52 precludes one. Regression discontinuity analysis is appropriate for this clear separation of eligible versus ineligible soldiers to become officers. The simplest and most commonly employed form of RD estimates two separate linear regressions, one for values below the critical score and another for values equal to and above the critical score. A third regressor permits a discrete gap or discontinuity between these two regression equations.

We estimate the following RD model:

$$outcome_i = \beta_0 + \beta_1 * kaba_i + \beta_2 * (kaba_i * kaba_i \text{ above } x) + \beta_3 * kaba_i \text{ above } x + \beta_4 * Z + \varepsilon_i, \quad (1)$$

where the dependent measure "outcome_{*i*}" refers to soldier *i*'s reported die outcome. The independent measures are: "kaba_{*i*}": soldier *i*'s military entrance score (minus *x* for ease of interpretation); "kaba_{*i*} above *x*": a dummy variable equal to 1 if soldier *i* scored *x* or more

different units require different *kaba* thresholds, a recruit's acceptance to or rejection from his preferred unit(s) provides him with an update about the possible range of his *kaba*.

on his entrance score and 0 otherwise; and " $kaba_i * kaba_i \text{ above } x$ ": an interaction term between these two variables. Z is a vector of control variables. According to this specification, β_1 captures the slope of the relationship between the reported outcome and the entrance score to the left of the discontinuity at x , $\beta_1 + \beta_2$ captures the slope of this relationship to the right of the discontinuity at x , while β_3 measures the magnitude of the discontinuity.

The difference between the RD models reported in the columns of Table 3 is that the critical entrance score, x , is allowed to vary. Regression (8) begins with the true critical score of 52. The estimates show no significant relationship between a soldier's $kaba$ and his reported die toss below nor above the score of 52. Neither the coefficient of -0.10 on $kaba_i$ nor the slope of the regression line above a $kaba$ of 52 ($\beta_1 + \beta_2 = .06$) is significantly different from 0 ($p = .15$ and $p = .52$, respectively). However, the weakly significant coefficient of -0.49 on the indicator variable $kaba_i \text{ above } x$ indicates that these two zero-sloped line segments are separated by a negative gap of 0.49 units. In other words, there is no difference in the reported die outcomes among soldiers with entrance scores below 52. Similarly, there is no difference in the reported outcomes among soldiers with scores of 52 and higher. However, there is a significant difference between these two groups: those with scores greater than or equal to 52 are more honest, reporting die outcomes half a point less than those with scores below 52.

<insert Table 4 about here>

When controls are added to regression (9), the above conclusions are further strengthened. Neither the slope to the left nor the right of the cutoff $kaba$ of 52 is significantly different from 0. Moreover, the gap between the two zero-sloped regression lines increases to -0.81 and is now highly significant ($p = .001$).

If the cutoff score of 52 has indeed been optimally set to classify soldiers according to their degree of honesty, then RD designs employing other placebo critical entrance scores ought to do less well at sorting the two groups by their levels of honesty. Regression (10) sets the $kaba$ cutoff at 50. Similar to the results in (8), the slopes to the left and to the right of the cutoff are not significantly different from 0. In (10), however, the discontinuity between these two lines is also not significant. Apparently, the inclusion of the less honest soldiers with entrance scores of 50 and 51 in the group above the cutoff

dilutes this group to the point that its soldiers are no longer more honest than those below the cutoff of 50. These same results obtain when controls are included in regression (11).

Regressions (12) and (13) ought to reduce this dilution effect by raising the entrance-score cutoff to 51. In (12), the discontinuity between the two groups of -0.42 points continues to be insignificant ($p=.17$), whereas the addition of control variables in (13) increases the magnitude of the discontinuity to $-.64$ to the point where it is now significant ($p=.05$). Still, both the magnitudes and degrees of significance of the discontinuity between the two groups are smaller in (12) and (13) compared to the corresponding regressions (8) and (9) in which the cutoff is set one point higher at 52. The implication is that a cutoff that excludes soldiers with a *kaba* of 51 leads to a more honest pool of soldiers than a cutoff that includes them.

Raising the entrance-score cutoff to 53 or 54 implies that soldiers with scores of 52 or 52 and 53, respectively, no longer qualify to be officers. With such a move, regressions (14) – (17) all show that neither the slope to the right of the cutoff nor the discontinuity at the cutoff is significantly different from 0. However, the slope to the left of the cutoff is now highly significant and equal to -0.21 and -0.20 when the cutoff equals 53 and -0.17 and -0.16 when the cutoff is 54. This finding implies that, when included in the group below the cutoff, soldiers with *kaba* scores of 52 and 53 (who are apparently more honest) stand out from the group in their reported die outcomes sufficiently to produce a significant negatively sloped relationship to the left of the cutoff.

To summarize, lowering the cutoff below the actual one of 52 would dilute the honesty of the pool of soldiers who qualify to be officers. At the same time, raising the cutoff above 52 would shrink the pool of eligible soldiers without any concurrent increase in the level of honesty. A cutoff of 52 maintains a sizeable pool of soldiers eligible to be officers without compromising on the level of honesty. To enhance further this pool, we presented evidence that the Israeli military avails itself of the subjective components of the entrance exam to boost deserving soldiers' scores to meet this cutoff.

4. Conclusions

The Israeli Defense Forces requires every entering soldier to take a series of psycho-technical exams and males to undergo an in-person interview. The military uses these

measures to assign each soldier to a military unit commensurate with his abilities and to determine whether he is eligible to become an officer in the military upon completion of his mandatory service. One of the paramount goals of these tests, including the interview, is to assess accurately the honesty of the incoming soldier.

The results from our experiment reveal not only a positive relationship between a soldier's test score and his honesty, but also that the Israeli military has optimally designed and implemented these series of tests to screen for honesty.

Companies for whom the honesty of their employees is of primordial importance would do well to devote more resources to adopt similar measures when hiring. In particular, the IDF employs two features that distinguish their screening procedure from those used by companies. First, far from being ordinary human resource personnel, the IDF interviewers are a select group of exceptionally intelligent – among the top 10% of the population – highly skilled individuals who have undergone specialized training in assessing others' personality traits. Second, soldiers endure an entire day of screening consisting of a series of written tests and an interview that often lasts up to two hours. This time commitment to assessing candidates is unparalleled. While similar commitments may be impractical and prohibitively expensive for most employers to apply to all job applicants invited for a first interview, they could be applied to the small number of short-listed candidates at the final stage of the selection process. What seems clear is that their ability to select upon honesty would represent an improvement over the current use of tricky, self-report honesty measures shown to be of dubious value.

References

Abeler, Johannes, Anke Becker and Armin Falk (2014) “Representative Evidence on Lying Costs,” *Journal of Public Economics*, 113, 96–104.

Association of Certified Fraud Examiners (1996) *Report to the Nation: Occupational Fraud and Abuse: 2014 Global Fraud Study*. Association of Certified Fraud Examiners, available at: <http://www.acfe.com/rtn/docs/2014-report-to-nations.pdf>.

Azar, Ofer. H., Shira Yosef and Michael Bar-Eli (2013) “Do customers return excessive change in a restaurant? A field experiment on dishonesty,” *Journal of Economic Behavior and Organization*, 93, 219–226.

Barrick, Murray R. and Michael K. Mount (1991) “Effects of impression management and self-deception on the predictive validity of personality constructs,” *Journal of Applied Psychology*, 81, 261–272.

Brooks, Chad (2013) “Employee theft on the rise and expected to get worse,” *Business News Daily*, June 19, 2013, available at <http://www.businessnewsdaily.com/4657-employee-theft-rising.html>.

Charness, Gary and Martin Dufwenberg (2006) “Promises and Partnerships,” *Econometrica*, 74:6, 1579–1601.

Dreber, Anna and Magnus Johannesson (2008) “Gender differences in deception,” *Economics Letters*, 99:1, 197–199.

Erat, Sanjiv and Uri Gneezy (2012) “White Lies,” *Management Science*, 58:4, 723-733.

Fischbacher, Urs and Franziska Föllmi-Heusi (2013) “Lies in disguise: An experimental study on cheating,” *Journal of the European Economic Association*, 11:3, 525–547.

Ghiselli, Edwin E. and Richard P. Barthol (1953) “The validity of personality inventories in the selection of employees,” *Journal of Applied Psychology*, 37:1, 18–20.

Guion, Robert M. and Richard F. Gottier (1965) “Validity of personality measures in personnel selection,” *Personnel Psychology*, 18, 135–164.

Gneezy, Uri (2005) “Deception: The role of consequences,” *American Economic Review*, 95:1, 384–394.

Hao, Li and Daniel Houser (2013) “Perceptions, Intentions, and Cheating,” unpublished manuscript.

Harrison, Glenn W. and John A. List (2004) “Field experiments,” *Journal of Economic Literature*, 42, 1009–1055.

Kahneman, Daniel (2002) “The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2002 Daniel Kahneman, Vernon L. Smith,” available at: http://www.nobelprize.org/nobel_prizes/economic-sciences/laureates/2002/kahneman-bio.html

Lahav, Eyal, Uri Benzion and Tal Shavit (2011) “The effect of military service on soldiers’ time preferences – Evidence from Israel,” *Judgment and Decision Making*, 6:2, 130–138.

Lee, David S. and Thomas Lemieux (2010) “Regression discontinuity designs in economics,” *Journal of Economic Literature*, 48:2, 281–355.

Lerer, Zeev (2009) "Groups of quality: The social history of the IDF selection system," Ph.D. dissertation, Tel Aviv University.

Levitt, Steven D. (2006) "White-collar crime writ small: A case study of bagels, donuts, and the honor system," *American Economic Review Papers and Proceedings*, 96:2, 290–294.

Ones, Deniz S., Stephan Dilchert, Chockalingam Viswesvaran and Timothy A. Judge (2007) "In support of personality assessment in organizational settings," *Personnel Psychology*, 60:4, 995–1027.

Pruckner, Gerald J. and Rupert Sausgruber (2013) "Honesty on the Streets: A Natural Field Experiment on Newspaper Purchasing," *Journal of the European Economic Association*, 1:3, 661–679.

Ruffle, Bradley J. and Yosef Tobol (2014) "Honest on Mondays: Honesty and the Temporal Distance between Decisions and Payoffs," *European Economic Review*, 65, 126–135.

Shalvi, Shaul, Jason Dana, Michael J.J. Handgraaf and Carsten K.W. De Dreu (2011) "Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior," *Organizational Behavior and Human Decision Processes*, 115, 181–190.

Tett, Robert P., Douglas N. Jackson and Mitchell Rothstein (1991) "Personality measures as predictors of job performance: A meta-analytic review," *Personnel Psychology*, 44, 703–742.

Wikipedia, [http://he.wikipedia.org/wiki/גיוס_לצה"ל](http://he.wikipedia.org/wiki/גיוס_לצה).

Table 1 – Descriptive statistics

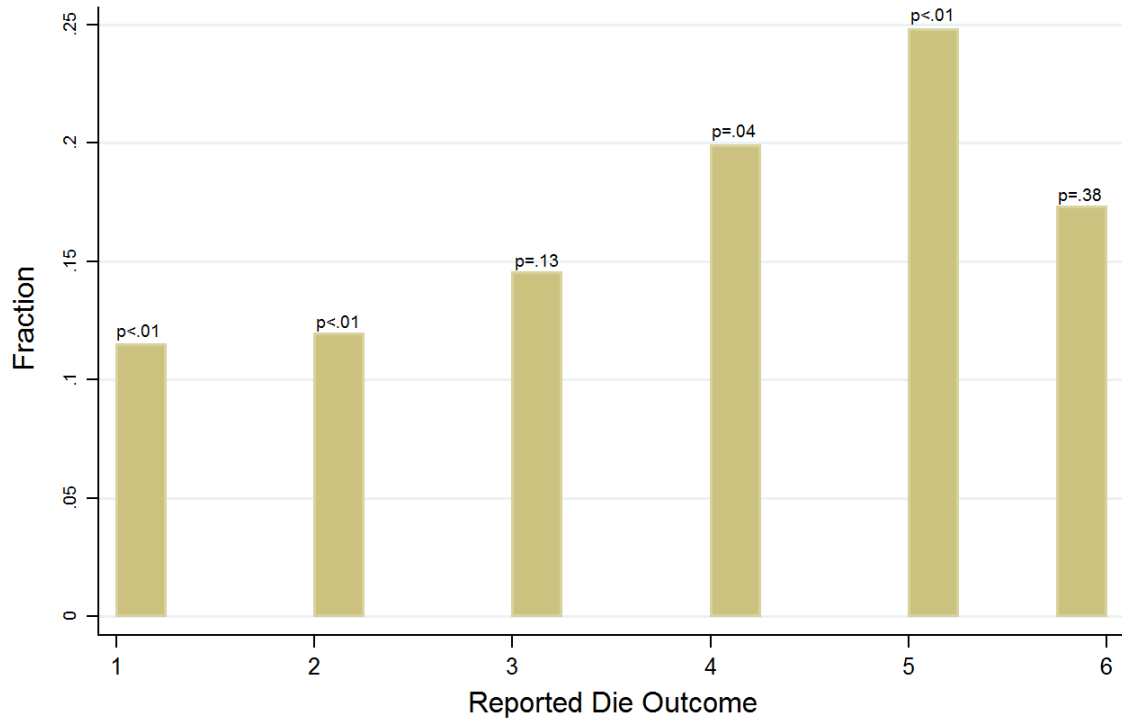
Variable (possible values)	Mean (s.d.)
<i>reported die outcome</i> (1,2,3,4,5,6)	3.87 (1.61)
<i>WTP for half hour early</i> (in Israeli NIS)	42.7, 30 (67.2)
<i>WTP for half hour early – outliers excluded</i> (in Israeli NIS)	33.9, 30 (28.1)
<i>WTP for 3 hours early</i> § (in Israeli NIS)	194.1, 100 (225.5)
<i>military entrance test score (kaba)</i> (45-56)	51.1 (2.5)
<i>claim know test score</i> § (0,1)	.839 (.369)
<i>actual know test score</i> § (0,1)	.406 (.492)
<i>female</i> (0,1)	.424 (.495)
<i>religious</i> (0,1)	.337 (.473)
<i>city resident</i> (0,1)	.724 (.458)
<i>self-reported honesty</i> (1=always tell truth – 4=truth when convenient)	2.12 (0.99)
<i>others think</i> § (1=very important – 7=not important at all)	3.68 (1.96)

Notes: 1) The die outcome refers to the roll of the die reported by soldiers; the WTP variables indicate the willingness to pay for a half-hour and a three-hour early release from the base (median values are displayed to the right of the means); the military test score outcomes in our sample range from 45-56; the indicator variables "*claim know test score*", "*actual know test score*", "*female*", "*religious*" and "*city resident*" indicate the fraction of soldiers that claims to know their army test score, the fraction that knows their true test score; the fractions of female soldiers, religious soldiers and soldiers from cities, respectively; participants also answered questions on the extent to which they generally tell the truth ("*self-reported honesty*") and how important it is to them what others think of them ("*others think*"). See the Appendix for the precise wording of the questions.

2) § indicates that the question appeared for the last 217 soldiers only. All other statistics are based on the full sample of 427 soldiers.

Figure 1

Histogram of Reported Die Outcomes



Note: p-value from one-sided binomial test that observed frequency of each die outcome is less (greater) than .166 appears above each bar.

Figure 2

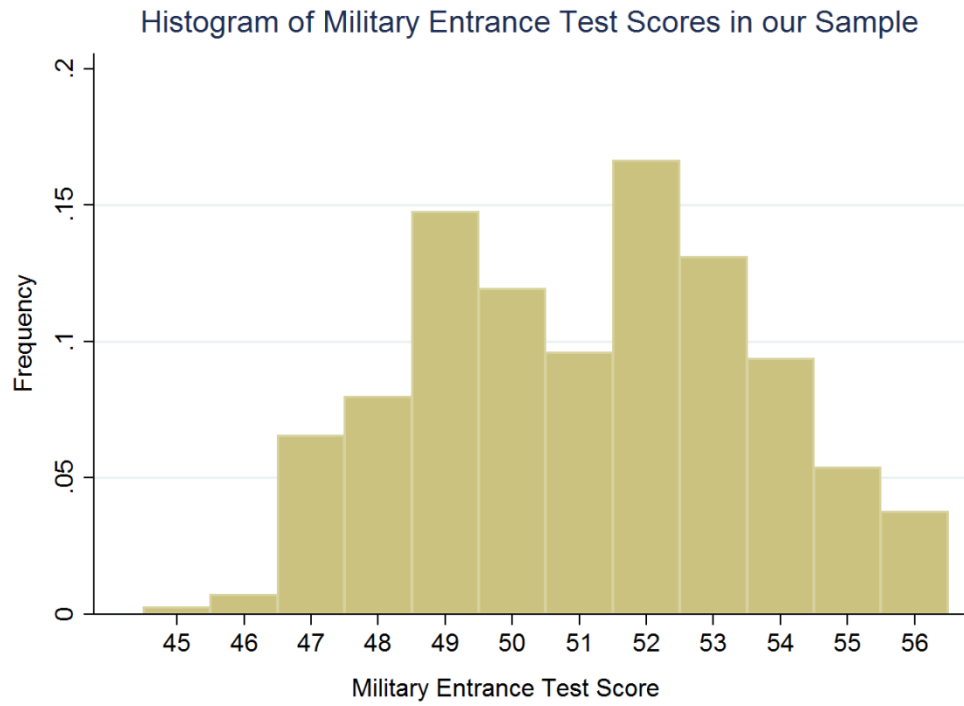


Figure 3

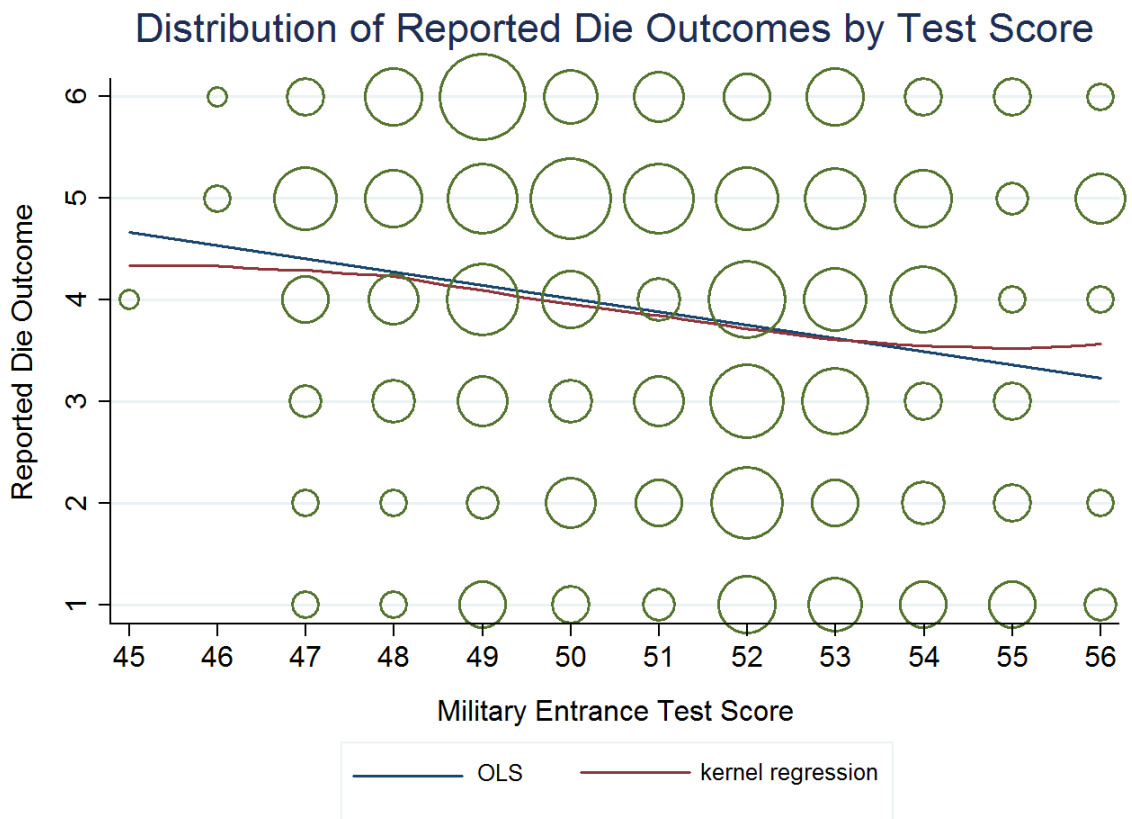


Table 2 – Regression analysis

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
constant	3.76*** (0.08)	1.92** (0.93)	2.06** (0.98)	3.78*** (0.16)	5.34*** (1.96)	3.67*** (0.13)	5.71*** (1.88)
Δ kaba from 52	-.13*** (.03)	-.10*** (.03)	-.11*** (.03)	—	—	-.18*** (.05)	-.13** (.06)
Δ kaba guess from 52	—	—	—	.09 (.08)	.05 (.08)	—	—
WTP for half hour early release	—	.001 (.001)	.007*** (.003)	—	.005 (.005)	—	.005 (.005)
female	—	.34* (.18)	.25 (.19)	—	.48 (.31)	—	.44 (.31)
religious	—	-.21 (.29)	-.15 (.31)	—	.27 (.62)	—	.27 (.62)
religious female	—	-.12 (.36)	-.16 (.37)	—	-.63 (.67)	—	-.62 (.67)
city resident	—	.37** (.17)	.30* (.17)	—	.54* (.29)	—	.49* (.29)
self-reported honesty	—	.06 (.08)	.08 (.08)	—	.30** (.14)	—	.25* (.14)
Adj. R ²	.04	.08	.10	.00	.15	.05	.17
Includes day-of-week, peer-effects controls	No	Yes	Yes	No	Yes	No	Yes
Excludes WTP Outliers	No	No	Yes	No	Yes	No	Yes
N	427	427	412	185	174	185	174

Notes: 1. Dependent variable: soldier i 's reported die outcome.

2. Regressors: Δ kaba from 52 is soldier i 's military test score (*kaba*) minus 52; Δ kaba guess from 52 is soldier i 's guess about his *kaba* minus 52; the soldier's willingness-to-pay for half an hour early release from the base; indicator variables for whether the soldier is female, religious, and an interaction term between the two, from a city (or a rural area); his self-reported honesty (question 4 in the Appendix), and measures of soldier i 's military unit and military base peer effects calculated as the mean reported die outcome of all members of soldier i 's unit and base, respectively, excluding soldier i .

3. Heteroskedasticity-robust standard errors in parentheses.

4. Regressions (2), (3), (5) and (7) include indicator variables for the day of the week on which soldier i participated in the experiment and measures of military unit and military base peer effects.

5. Regressions (3), (5) and (7) exclude observations more than two standard deviations above the mean “WTP for half hour early release”.

6. Coefficient significantly different from 0 at the 1% level ***, at the 5% level **, at the 10% level *.

Table 3 – Regression discontinuity models

Variable	Coeff.	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
<i>Constant</i>	β_0	3.91 (0.24)	2.31 (1.01)	4.32 (0.34)	2.63 (1.00)	4.06 (0.25)	2.61 (0.99)	3.34 (0.18)	1.79 (0.96)	3.30 (0.18)	1.75 (0.95)
Δ <i>kaba from x</i>	β_1	-.10 (.07)	-.03 (.08)	-.01 (.13)	.04 (.14)	-.09 (.09)	-.01 (.10)	-.21*** (.05)	-.20*** (.05)	-.17*** (.04)	-.16*** (.04)
Δ <i>kaba from x</i> * <i>kaba above x</i>	β_2	.16 (.12)	.18 (.12)	-.08 (.15)	-.11 (.16)	.05 (.12)	.02 (.13)	.18 (.15)	.26 (.16)	.22 (.25)	.35 (.25)
<i>kaba above x</i>	β_3	-.49* (.28)	-.81*** (.30)	-.45 (.33)	-.45 (.34)	-.42 (.30)	-.64** (.32)	.27 (.27)	.25 (.27)	.21 (.30)	.19 (.29)
slope above x	$\beta_1+\beta_2$.06 (.09)	.15 (.09)	-.09 (.06)	-.07 (.06)	-.04 (.08)	.01 (.08)	-.02 (.14)	.06 (.15)	.04 (.25)	.19 (.24)
<i>x (kaba threshold)</i>		52	52	50	50	51	51	53	53	54	54
Controls		No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R ²		.07	.12	.04	.10	.04	.11	.04	.11	.04	.10
N		427	412	427	412	427	412	427	412	427	412

Notes: 1. Dependent variable: soldier *i*'s reported die outcome.

2. Regression discontinuity models with the existing *kaba* cutoff of 52 that qualifies one to be a soldier (8)-(9) and placebo cutoffs below the existing cutoff ((10)-(13)) and above the cutoff ((14)-(17)). The last row before the control variables, *x*, indicates the *kaba* threshold under consideration.

3. The column "Coeff." refers to the regression coefficients specified in the regression model of equation (1) in the text. Explanations of the independent measures appear beneath the equation in the text.

4. Heteroskedasticity-robust standard errors in parentheses.

5. Coefficient significantly different from 0 at the 1% level ***, at the 5% level **, at the 10% level *.

6. The set of controls included in the odd-numbered regressions are: WTP for half hour early release, military base and unit peer effects, self-reported honesty and dummy variables for female, religious, their interaction, city residents and the day of the week the experiment took place. Table 1 provides an explanation of the control variables.

Appendix – Questionnaire

Note: § refers to questions introduced after 210 soldiers had already participated.

1. What is the maximum amount of money you would be willing to pay to be released to go home half an hour earlier than the scheduled time on your day of release?

_____ (amount in NIS)

§ 2. What is the maximum amount of money you would be willing to pay to be released to go home three hours earlier than the scheduled time on your day of release?

_____ (amount in NIS)

3. How would you define yourself?

1. secular
2. traditional
3. religious
4. ultra-orthodox

4. Which of the following sentences best describes you?

1. I always tell the truth.
2. I almost always tell the truth.
3. I usually tell the truth.
4. I tell the truth when it is convenient for me.

§ 5. How important is it to you what others think of you?

1=very important, 7=not important at all

1 2 3 4 5 6 7

6. Where do you live (indicate the name of the town or city)? _____

§ 7a. Do you know your army test score?

0. no
1. yes

§ 7b. If yes, please write it in the space provided. _____

Thank you for your participation.