DISCUSSION PAPER SERIES

DP12208

SELF-WORTH VERSUS NET WORTH: IMAGE MOTIVATION AND THE QUANTITY-QUALITY TRADE-OFF

Michelle Brock

DEVELOPMENT ECONOMICS

ISSN 0265-8003

SELF-WORTH VERSUS NET WORTH: IMAGE MOTIVATION AND THE QUANTITY-QUALITY TRADE-OFF

Michelle Brock

Discussion Paper DP12208
Published 07 August 2017
Submitted 07 August 2017

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SELF-WORTH VERSUS NET WORTH: IMAGE MOTIVATION AND THE QUANTITY-QUALITY TRADE-OFF

Abstract

Self-image concerns can motivate behaviour over which it is difficult to contract. We investigate how self-image concerns impact the quantity-quality trade-off of professionals, independent of external reputation. To do so, we use a framed field experiment with judges. The experiment measures the effect of self-image concerns compared to that of a bonus and to that of combing both incentives. We find that while both incentives increase the quantity of work, bonuses generate much worse quality. Subjects are less willing to engage in opportunistic behaviour to earn the bonus, however, if a peer might see their work, albeit anonymously.

JEL Classification: D91, M52, J33

Keywords: Self-image, judges, professionalism, real-effort task, intrinsic incentives, Tajikistan

Michelle Brock - brockm@ebrd.com
EBRD and CEPR

Acknowledgements
This work was funded by a Technical Cooperation grant from the European Bank for Reconstruction and Development. We are grateful for the support of the Judicial Training Center and the Council of Justice in Tajikistan. We acknowledge excellent research assistance from Carly Petracco and Muzaffar Ahunov. We benefited from comments from Catherine Eckel, Sergei Guriev, R'oi Sultan, Nathanial Young, EBRD OCE Seminar Series participants, and participants at several ESA conferences.

Self-worth versus net worth: Image motivation and the quantity-quality trade-off

J. Michelle Brock

Abstract

Self-image concerns can motivate behaviour over which it is difficult to contract. We investigate how self-image concerns impact the quantity-quality trade-off of professionals, independent of external reputation. To do so, we use a framed field experiment with judges. The experiment measures the effect of self-image concerns compared to that of a bonus and to that of combing both incentives. We find that while both incentives increase the quantity of work, bonuses generate much worse quality. Subjects are less willing to engage in opportunistic behavior to earn the bonus, however, if a peer might see their work, albeit anonymously.

Keywords: self-image, judges, professionalism, real-effort task, intrinsic incentives, Tajikistan.
JEL Classification: D91, M52, J33

Contact details: European Bank for Reconstruction and Development and CEPR
One Exchange Square, London, EC2A 2JN, *Phone*: +44 2073387193; *Fax*: +44 2073386111;
Email: brockm@ebrd.com.

This work was funded by a Technical Cooperation grant from the European Bank for Reconstruction and Development. We are grateful for the support of the Judicial Training Center and the Council of Justice in Tajikistan. We acknowledge excellent research assistance from Carly Petracco and Muzaffar Ahunov. We benefited from comments from Catherine Eckel, Sergei Guriev, R'oi Sultan, Nathanial Young, EBRD OCE Seminar Series participants, and participants at several ESA conferences.

1 Introduction

According to the classic theory on multi-tasking, workers will put as little effort as possible into their jobs and only along the lines of what is in their contract. Effort is limited and workers will allocate effort in order to earn rewards (Holmstrom and Milgrom, 1991). Moreover, while quantity may be contracted upon, it is difficult to incentivize quantity without compromising quality. But despite incomplete contracts, professionals – judges, lawyers, teachers, doctors and others – regularly deliver on quality. This suggests additional rewards to effort beyond the explicit contractual ones.

In the public service professions mentioned, image rewards from acting according to professional norms can be important non-contractual incentives for motivating quality (Akerlof and Kranton, 2000). For example, self-image at work will be heavily influenced by professional norms, which provide benchmarks for performance quality.[1] When individuals internalize a norm, they will adhere to it as a way to maintain cognitive consistency with respect to their own self-image (Ryan and Deci, 2000). Highly skilled and specifically trained groups, such as the judiciary, teachers and doctors get exposure to professional norms during their training and may even belong to professional associations, which further propagate expectations of behavior. Thus, people in these groups may get utility from norm conformance or disutility from not complying with norms.

Obtaining utility from conforming to norms is a type of social motivation. And

[1]Professional norms can be thought of as agreed upon codes of conduct that drive decision making among the individuals in a group of common expertise, training and occupation.

when social motivation is strong, explicit incentives can be counterproductive. For example, performing a charitable act can be a signal to oneself and others about deeply held values. If a charity offers to pay volunteers, the signal gets muddled – it becomes unclear if the volunteer is acting due to deeply held values or in order to earn money – and effort goes down (Bowles and Polania-Reyes, 2012). The same logic can be applied to professions with a social service dimension. It can be complicated, then, to design incentive schemes for such workers that reward quality without undermining social motivation.

The contribution of this paper is to show how monetary and image incentives compare and interact in a professional setting with multi-tasking. We use peer scrutiny to trigger image concerns. Peer scrutiny is a powerful example of an image incentive that can influence a professional's quantity-quality trade-off. When we know our work our will be scrutinized by our professional peers, we work harder to make sure it will meet the group's standards (Akerlof and Kranton, 2008). In doing so, we reinforce our commitment to the group values as well as our own professional identity, which has implications for our utility optimization. Peer scrutiny and the corresponding image rewards thus leverage professional self-image to influence workplace behavior.

For the study we used a framed field experiment. The treatments vary the visibility of work performance among peers (i.e. peer scrutiny), presence of a competitive bonus and the interaction of the two (increased visibility and bonus together). We designed the visibility treatment – a one shot setting with no feedback and where the audience is uncertain – such that the gains from socially desirable behavior are the

result of an internal feedback process where the individual acts to conform to norms and thus understands that he has social approval, were it to be explicitly given. We chose a competitive bonus as the high-powered monetary incentive because we expect it to act primarily on volume; a backlog of cases has been identified as an important performance failure in our sample.

We explore these issues using a sample that particularly lends itself to our research aim, namely judges in Tajikistan. The value of courts for an economy depends in part on how judicial professionals make trade-offs between quantity and quality. Institutional elements that reinforce quality-based reputation or career concerns for judges are absent in Tajikistan, making it an ideal place to look at intrinsic motivation to perform. Preferences for quality that stem from intrinsic sources, like self-image, may be instrumental for maintaining a minimum level of quality in the system.

The role of social rewards for judges' performance, and the performance of other social service professionals, has important social and economic implications. Judges are responsible for upholding legal institutions and their decisions impact firms' ability to thrive. On-going debates about performance in healthcare and academia center around the question of how to achieve a desired volume of output – patients treated or research papers published – without compromising quality. Primary school teachers perform an essential task for society and at the same time receive modest incomes (less in the developing world). Image incentives likely play a large role in motivating performance in these professions. As Bowles and Polania-Reyes (2012) suggest, leveraging such socially linked incentives effectively can not only reduce crowding

out of intrinsic motivation, but can also increase complementarity with monetary incentives.

Our results confirm the importance of visibility in how judges allocate effort to quantity and quality. We find that while visibility increases effort on both dimensions, the portion of total effort put toward quality is no higher with visibility than in the control. A monetary incentive in the form of a bonus increases effort only on quantity, with 53% of participants choosing to put less into quality than quantity. Lastly, combining a bonus with visibility is more effective than the bonus alone for improving quality. The results are consistent with a model with self-image concerns, over both quantity and quality, that balance out motivation from pure financial incentives. Importantly, our combined bonus and visibility treatment provides evidence that self-image can crowd in socially desirable behavior, even when such behavior will ultimately lead to higher earnings.

The paper proceeds as follows. In section 2 we discuss the literature. In section 3 we present a model of how professionals determine quality and quantity in the presence of concerns for self-image. Section 4 provides the details of the study, including descriptions of: the Tajik judicial system, the participant sample, the experimental design and the empirical specification. Results appear in section 5 and a final section concludes.

2 Literature review

This paper contributes to the literature on crowding out of intrinsic motivation. The question in the crowding out literature is whether monetary incentives to "do well" crowd out non-monetary incentives to "do good" (Ariely, Bracha and Meier, 2009). Money's effectiveness to motivate effort is legitimately questioned because, while it acts directly as an incentive to perform a task, it can also act indirectly to undermine intrinsic motivation. In particular, monetary incentives can remove expectations of moral or socially motivated behavior, triggering "moral disengagement" (Bowles and Polania-Reyes, 2012; Gneezy and Rustichini, 2000).

While the bulk of the empirical literature on crowding out examines effort expended doing charitable activities such as volunteering or donating blood, a few papers have also expanded this research to a workplace setting. Monetary incentives may crowd out intrinsic motivation of a worker, the same way they do in the case of charitable actions. This can happen for a variety of reasons: money may act as a signal from principals to agents about the quality of a task, principals may offer money as a result of distrust between themselves and the agent, or money may reduce the image gains behaving like a professional (Frey and Oberholzer-Gee, 1997; Gneezy, Meier and Rey-Biel, 2011).

Ariely, Bracha and Meier (2009) provide evidence of how image concerns can lead to crowding out in a real-effort task. In their experiment, students worked to raise money for charities. The authors also randomized visibility of the actions, and only a random selection of subjects got to earned money for themselves. Because

doing charitable work for money is seen as socially undesirable behavior, image gains come from *not* working harder for money. The authors find that monetary incentives can increase effort on pro-social tasks, but visibility reduces the benefit to the agent from responding to them.

This result is echoed in Carpenter and Myers (2010)'s study on volunteer firefighters, which also shows that visibility matters disproportionately for those with stronger image concerns. Previous studies have simply assumed this connection. Presence of image concerns is captured with whether or not a firefighter had a vanity license plate. Carpenter and co-authors find that those with image concerns are more likely to show up on a call, "supporting predictions that the effect of image concerns increases with the visibility of the activity". Carpenter et al. does not look at the quantity quality trade-off and while volunteer fire-fighters have a public sector role, they are not public sector employees.

Our paper adds to this literature by bringing public sector professionals into the laboratory. This is unique and important because public sector professionals face a different set of non-contractual motivations from volunteers, student workers or private sector professionals. Moreover, we study the differential impact of the monetary and non-monetary incentives in a multi-tasking environment.

This study also contributes to the growing literature on image rewards and non-monetary incentives. Image concerns have been shown to be influential for both pro-social behavior and worker effort choices. Researchers study image concerns by manipulating visibility of actions. For example, Ashraf, Bandiera and Jack (2014) and

Kosfeld and Neckermann (2011) use increased visibility as a treatment in their laboratory and field experiments, respectively. In a non-experimental paper, Lacetera and Macis (2010) show how public recognition with small trophies increases blood donations in a small town. These papers have a particular focus on the status conferred by publicly announcing top performers. The consensus from this literature is that image concerns matter; these public announcements can increase performance by up to 100%.

The key feature of the visibility treatments in these studies is that some aspect of performance gets publicized and that, unlike this paper, the publicity is not anonymous. Also, the experimental evidence so far has been collected using student samples. For example, in the Ariely et al. visibility treatment, every subject had to announce their performance in front of other participants after the experimental tasks were completed. In Kosfeld and Neckermann (2011), students recruited for a data entry job competed for a publicaly awarded non-monetary prize (i.e. a congratulatory certificate). The students in this visibility treatment had 12 per cent higher performance than workers in a treatment without the reward. In both studies, all subjects are exposed to the prospect of publicity, but subjects must compete for it and not everyone has the same probability of winning. Also, neither of these studies separate out the impact of competition, per se, from the impact of the visibility incentive. Thus, while the results clearly demonstrate the power of public recognition, they are more precisely about using public recognition and social comparison together to motivate workers.

Moreover, since these studies do not maintain anonymity of subjects, the treat-

ment effect includes influences from both the increased visibility as well as identifiability of the workers. Determining the relative importance of each influence matters because non-anonymous visibility policies will not work well in many settings. Non-anonymous publicity may work well in small settings, where the audience is composed of the same people that competed for the prize. But when the audience is larger, each individual may be effectively anonymous. Also, the prize itself may be unimportant on average if the portion of the population that competed for the prize is small. Lastly, and perhaps most importantly, is that non-anonymous recognition may not be feasible at workplaces where incomes and rankings are not disclosed.

A few other studies besides ours preserve anonymity of subjects. This is usually done by sharing productivity distributions (histograms) publicly, while individual rank in the distribution is private information (Bandiera, Barankay and Rasul, 2013; Charness, Masclet and Villeval, 2013; DellaVigna and Pope, 2016). In these studies, the authors find that publishing the performance and rankings in the group increases performance, compared to the control. But these effects cannot be due to visibility alone, since the relative performance information can also induce more competitive behavior. Our paper will maintain anonymity and all participants will face the same pressure from having their work observable by their peers, but ranking information will not be available.

Using visibility to generate utility gains from self-image is based on self-perception theory. According to self-perception theory, stimuli motivate actions, which in turn determine and/or reinforce the beliefs we have about ourselves (Bem, 1972). In this study, the peer scrutiny is the stimulus, the response to that is the quantity and qual-

ity effort choice and the resulting belief is the professional identity supported by that action.

Providing a stimulus for self-image investment can be especially important in the presence of competing incentives or incompletely internalized professional norms (Ryan and Deci, 2000). Competing incentives may include extrinsic rewards or other norms. Individuals may even have competing elements of their self-image between which they make trade-offs when deciding behavior. The term 'identity salience' describes the fact that at different points in time or in different contexts, an individual's behavior may be more or less influenced by any one identity (Turner and Onorato, 1999). Identity salience can explain why one may not act according to professional norms in the absence of external stimulus or a reminder to do so (Cohn, Fehr and Maréchal, 2014; Hoff and Pandey, 2014). In this case, the identity of "the person who likes to conform to professional norms for quality" may not be a salient identity absent a stimulus. Peer scrutiny can serve as such a stimulus, or reminder, to conform to professional norms. Self-image gains from norm conformance will come from reflecting on having behaved in a way that is consistent with external expectations. Increasing opportunities for such rewards may thus increase compliance with professional norms for quality.

This theoretical approach contrasts somewhat with the framing theory of crowding out. The framing theory hypothesizes that incentive structures frame exchanges: providing either monetary or non-monetary incentives indicates to people how to behave, and how their own behavior may be interpreted by others (Heyman and Ariely, 2004; New and Titmuss, 1972). Non-monetary incentives can be anything

from social approval to symbolic rewards to small gifts. Money reduces pro-social behavior because intrinsic motives are very active in the charity frame and less active in the money frame. An alternative view of crowding out has been offered by Seabright (2009). He argues that whether or not monetary incentives work to motivate behavior comes down to the value of the signal each behavior would send to others, and not the framing of the task, per se. The signal is important because it determines the types of people in one's network. For example, it is valuable to be in a network with people who reciprocate charitable acts, such as a helping one move or helping with childcare, without the expectation of being paid. A framework based in self-perception theory is more consistent with Seabright than with New and Titmuss, because the signal plays a central role in determining behavior. Also, while a workplace is always in a monetary frame, non-monetary and social incentives do play a role in determining behavior. Thus, to think holistically about crowding out in the workplace, one must go beyond the framing explanation.

3 A model of worker effort and self-image utility

To clarify ideas, we specify a model where adhering to professional norms delivers rewards to a decision-maker in the form of self-image utility (as in Bénabou and Tirole (2003) and Bowles and Polania-Reyes (2012)). In our model, an agent choses effort to maximize their utility from work. Effort is revealed through how much work is done and how much of that work is done well. We represent total work completed with w and the amount out of the total that is done well with q, where $q \leq w$. For

example, if a judge passes down rulings on 10 cases in a month and 3 of them are poor quality, then $w = 10$ and $q = 7$. q and w can also pertain to sub-components of a single project. Consider, for example, a lawyer who must do a number of tasks when they prepare a case. Suppose a lawyer's tasks include applying the law correctly, compiling evidence, producing written arguments and interfacing with the parties involved in the case. If they do only two of these well, $w = 4$ and $q = 2$. In this setting, the socially desirable behavior is to limit the volume and fraction of low quality work, $(w - q)$ and $\frac{w-q}{w}$, respectively. Having a higher total output (with quality non-decreasing), $(w|\frac{w-q}{w})$, may also be considered socially desirable, if speed of justice is a major problem. But improved speed without at least constant quality is not desirable.

Total utility from work is a linear combination of returns to effort, $Y(w,q)$, self-image gains, $R(w,q)$, and effort cost, $C(w,q)$.

$$U(w,q) = Y(w,q) + R(w,q) - C(w,q) \tag{1}$$

We specify the model further following Bénabou and Tirole (2006), but with modifications to allow analysis of the quantity-quality trade-off in a professional setting. The model describes the decision making of a worker who must chose effort to complete a series of tasks. Effort is a pair (w,q). Income, $Y(w,q)$, will include a rank-order tournament for winning a bonus payment. Cost of effort is $C(w,q) = \kappa w^2/2 + \varepsilon q^2/2$. Propositions about the treatment effects in the experiment follow from differences in the derived expression for (w^*, q^*) when parameters for visibility and bonus are alternately set to zero.

$Y(w,q)$ includes both monetary and intrinsic returns to effort. It is a linear combination of quantity and quality, where intrinsic valuations appear as weights on total earnings from each kind of effort. Workers earn p per completed task and an additional g if the task is completed well (i.e. the employer pays a piece-rate premium for quality). Intrinsic returns to effort that the worker puts into quantity and quality are denoted v_w and v_q, respectively. Workers can also earn a bonus if they are a top performer, where performance is measured as the total value of the tasks completed, $pw+gq$.[2] We include the bonus option in the model because it is a common feature of employee compensation (Lazear and Rosen, 1981), which we also use in the experiment. B is the value of the bonus. If the bonus is unavailable, $B=0$. The employee's expected value of winning a bonus is $E[win] = B * \rho[(pw+gq)_i \geq (pw+gq)_{-i}]$, where ρ is a weight less than or equal to 1. Income is then expressed as:

$$Y(w,q) = v_w pw + v_q gq + E[win] \tag{2}$$

where p is the per unit wage for completing an item and g is the additional reward for completing the item well. Total gains for completing an item well are then $p+g$. Since performance is an aggregate measure in this model, an employee can win the bonus from choosing any convex combination of quantity and quality, conditional on their total productivity being greater than that of their peers. $E[win]$ can thus be broken down further to reflect the fact that people may care about how they win a

[2] In many social service professions, income is a more a function of tenure than performance. In our study we allow income to be a function of performance value because, even in the public sector, managers often have leeway to set income and bonuses according to performance, albeit in fixed ranges.

bonus, with either more points from q or more from $(w-q)$. We discuss this in more detail, below, with the description of self-image utility.

Self-image utility, $R(w,q)$, consists of image gains from doing more and from doing well.[3] We specify it as a linear combination of returns to appearing skilled or appearing opportunistic, $E(v_q|q,g)$ and $E(v_w|w,p)$, respectively. $R(w,q)$'s importance in the utility functions depends on the visibility or salience of the effort choice, denoted by x.

$$R(w,q) = x\left[\gamma_w E(v_w|w,p) + \gamma_q E(v_q|q,g)\right] \tag{3}$$

with $\gamma_q \geq 0$. As in Bénabou and Tirole (2006) we let $\mathbf{v} = (v_w, v_q) \in \mathbb{R}^2$, where (v_w, v_q) is drawn independently from a continuous distribution with known mean and variance.

$$\begin{pmatrix} v_w \\ v_q \end{pmatrix} \sim N\left(\begin{matrix} \bar{v}_w \\ \bar{v}_q \end{matrix}, \begin{bmatrix} \sigma_w^2 & \sigma_{wq}^2 \\ \sigma_q^2 & \sigma_{wq}^2 \end{bmatrix}\right)$$

$E(v_q|q,g)$ and $E(v_w|w,p)$ are the worker's own posterior expectations of his own type $\mathbf{v}$. Realized type is private information. As in self-perception theory (Bem, 1972; Bodner and Prelec, 2003) we assume that an individual's characteristic, or

[3]The model can be expanded to include reputation, time spent at work and time spent at leisure, as in Posner (1993). We exclude external reputation from the model to more accurately reflect the situation in the field: self-image is potentially more important than signalling to others when visibility, and thus gains from reputation, is minimal or uncertain. Correspondingly, in the experiment all work is anonymous, to ensure that there are no opportunities to obtain returns to reputation. Also, our experiment is a one-shot game with no labor-leisure trade-offs, so this is also excluded from the model.

type, $\mathbf{v}$, is revealed to the individual himself through his own actions, (w,q), given the rate of monetary return (p,g). The individual then chooses his action to support the belief of the type he wishes to be - he updates his self-image distribution from $p(\mathbf{v})$ to $p(\mathbf{v}|w,q;p,g)$. As in Bodner and Prelec (2003) we consider one's type as given to the individual by the relevant social group. In this paper, that group is other judges.[4]

We now turn back to the idea that whether one wins the bonus by amassing points from quality versus from quantity matters for self-image utility. Winning a bonus from having a higher quality than others shows a commitment to quality norms, and thus would contribute positively to self-image. Preferences for a quality-based win will be included in the gains to self-image term $E(v_q|q,g)$. Winning a bonus from opportunistic behavior (i.e. completing many many items albeit with poor quality) is considered against professional norms and thus contributes negatively to self-image. Utility (or disutility) from winning a bonus by having a large w will be reflected in $E(v_w|w,p)$. Note that if the worker's productivity has no exposure to peers (i.e. $x=0$), the worker is not self-conscious over how he or she wins a bonus.

Note that self-image is not a function of reputation. It is the result of an internal assessment of one's type given one's own behaviour (Bem, 1972). γ_w and γ_q are similar to Bénabou and Tirole (2006)'s weights reflecting "the idea that people would like to appear as *prosocial* (public spirited) and *disinterested* (not greedy)", but in

[4]Bodner and Prelec (2003) also include a diagnostic utility in their model: total utility = outcome utility + diagnostic utility = $u(x,\theta)+\sum_{\theta} p(\theta|x)V(\theta)$, where $p(\theta)$ is the uncertainty of person *we* being type θ, conditional on choosing x. Also related is James Andreoni and B Douglas Bernheim (2009)'s signalling model. In their model the agent cares about *others'* perception of his type rather than his own perception of his type. In this case, the agent acts fairly to signal his type to the other.

this application each individual is their own judge of how an action makes them appear. The separable term for self-image utility is also similar to "psychological costs of cheating" in Nagin et al. (2002)'s rational cheater model. The psychological costs of cheating, $\chi(c)$, is a constraint on opportunism. Unlike our model, however, $\chi(c)$ does not include psychological gain from success. Also, we explicitly vary the salience of the psychological constraints and consider how this impacts the trade-offs workers make.

The first order conditions for this model give:

$$w^* = \frac{1}{\kappa}\left(v_w p + \frac{B\partial E[win]}{\partial w} + x\gamma_w \frac{\partial E(v_w|w,p)}{\partial w}\right) \tag{4}$$

$$q^* = \frac{1}{\varepsilon}\left(v_q + \frac{B\partial E[win]}{\partial q} + x\gamma_q \frac{\partial E(v_q|q,g)}{\partial q}\right) \tag{5}$$

To simplify the notation, let $E[win]_w = \frac{\partial E[win]}{\partial w}$ and $E[win]_q = \frac{\partial E[win]}{\partial w}$. Also, moving forward we will let q be the numeraire and all formulas will be expressed in terms of p.

To obtain clear predictions for the differences in performance across treatments, we calculate expressions for $E(v_w|w,p)$ and $E(v_q|q,g)$ using Bayesian updating results for Normal random variables (i.e. Kalman filter).

$$E[v_w|w,p] = \overline{v}_w + \frac{p^2\sigma_w^2 + Bcov(v_w, E[win]_w)}{p^2\sigma_w^2 + B^2 var(E[win]_w) + 2Bpcov(v_w, E[win]_w)}(w - E(w))\kappa \tag{6}$$

and

$$E[v_q|q,g] = \bar{v}_q + \frac{\sigma_q^2 + Bcov(v_q, E[win]_q)}{\sigma_q^2 + B^2 var(E[win]_q) + 2Bcov(v_q, E[win]_q)}(q - E(q))\varepsilon \quad (7)$$

For this paper, we look at what happens when we vary x and B. In the control group, there is no bonus and no visibility, so $x = 0$ and $B = 0$. This gives $w^* = \frac{v_w p}{\kappa}$ and $q^* = \frac{v_q}{\varepsilon}$. Workers will produce (w^*, q^*) such that marginal cost equals marginal benefit, where marginal benefit includes both extrinsic (monetary) and intrinsic returns. If $\mathbf{v} < 1$ then intrinsic returns to effort are a drag on production. If instead $\mathbf{v} \geq 1$ then intrinsic returns have either a neutral or positive impact on equilibrium output.

For the *visibility* treatment, where $x = 1$ and $B = 0$, the worker's self-image concerns are triggered. The optimization now includes returns to self image such that:

$$w^* = \frac{v_w p}{\kappa} + \gamma_w \kappa \qquad \text{and} \qquad q^* = \frac{v_q}{\varepsilon} + \gamma_q \varepsilon. \quad (8)$$

Here, $\gamma_q \varepsilon$ and $\gamma_w \kappa$ capture returns to appearing public spirited or greedy, respectively. Thus, according to the model, we expect that increases in visibility ($x > 0$) will impact the optimal (q^*, w^*) in so far as γ_q and γ_w are non-zero. As in Bénabou and Tirole (2006), we assume γ_q will be positive. This gives our first hypothesis.

Proposition 1 *Given $\gamma_q > 0$, q^* will be higher in the* visibility *treatment than it is in*

the control.

If q^* increases, w^* will also increase only if low quality work is at least at the same level as it is in the control. This will occur if $\gamma_w \geq 0$, as is seen in (8). But we do not take the sign of γ_w as given. This is because we are looking at image gains from worker effort, rather than from altruistic actions. Due to the public service nature of the profession, it may pay to appear disinterested, such that $\gamma_w < 0$. On the other hand, earning a high income may be image-positive, even at the expense of quality, such that $\gamma_w > 0$. Moreover, any increased effort aimed at doing more correctly may also result in accidentally also doing more wrong. For example, under time constraints, an intrinsically motivated worker will do their best but will not want to perform too many checks on their work because it can waste time. Consequently, increasing the salience of image concerns with visibility can lead to an acceptable increase in w. It is also not clear a priori whether a motive to appear disinterested would outweigh self-image gains to earning a lot of money from working. Given this complexity, we leave the impact of visibility on w^* as an empirical question.

We next discuss the effect of the bonus on behavior. In *bonus*, $B > 0$ such that $w^* = \frac{1}{\kappa}(v_w p + BE[win]_w)$ and $q^* = \frac{1}{\varepsilon}\left(v_q + BE[win]_q\right)$. Workers have an extra marginal benefit to produce in both dimensions, since it is possible to win B with high output of both q and $(w-q)$. However, the marginal gain to effort from winning the bonus is weighted by expectations of winning. Therefore, the difference in the optimal (w^*, q^*) between the bonus treatment and the control will be driven by the workers' expectations of winning from different kinds of effort. Workers who expect that they can produce enough quality answers to beat others' productivity, will

focus effort on producing q, while those who want to compete by completing many (incorrect) items quickly will focus on w without concern for q. Other workers may take a mixed strategy approach. Due to this potential heterogeneity and the fact that correct answers are worth twice as much as incorrect answers, we predict increases in w^* from both q^* and $(w^* - q^*)$, compared with the control. In addition, as per classic theory on piece-rate incentives, we expect that a bonus for total productivity will act as an indirect incentive for subjects to maximize quantity over quality, such that $E[win]_w > E[win]_q$ for most workers, and thus expect disproportionately large increases in $(w - q)$ for this treatment.

Proposition 2 *In the absence of visibility, a tournament bonus incentive will lead to increases in both q^* and w^*, with larger increases in $(w^* - q^*)$ than in q^*, compared with the* control.

We do not expect decreases in effort for this treatment, because we assume that $\frac{\partial E[win]}{\partial q} \geq 0$ and $\frac{\partial E[win]}{\partial w} \geq 0$ (i.e. the change in the probability of winning given an additional unit of effort cannot be negative).

In turn, the impact of *visibility+bonus* on w^* and q^* will depend on γ_w and γ_q, as well as how clear a signal behavior is (σ_w and σ_q) and the strength of the relationship between one's type, $\mathbf{v}$, and the marginal change in expectation of winning from q or from w ($E[win]_q$ and $E[win]_w$, respectively). In this treatment, both $B > 0$ and $x > 0$, such that:

$$w^* = \frac{1}{\kappa}\left(v_w p + BE[win]_w\right) + \gamma_w \Psi(w) \tag{9}$$

$$q^* = \frac{1}{\varepsilon}\left(v_q + BE[win]_q\right) + \gamma_q \Omega(q) \tag{10}$$

where

$$\Psi(w) = \left(\frac{p^2\sigma_w^2 + Bcov(v_w, E[win]_w)}{p^2\sigma_w^2 + B^2 var(E[win]_w) + 2Bpcov(v_w, E[win]_w)}\right)\kappa \tag{11}$$

and

$$\Omega(q) = \left(\frac{\sigma_q^2 + Bcov(v_q, E[win]_q)}{\sigma_q^2 + B^2 var(E[win]_q) + 2Bcov(v_q, E[win]_q)}\right)\varepsilon \tag{12}$$

Formulas (11) and (12) show that $\Psi(w)$ and $\Omega(q)$ will both be between -1 and 1; the distribution of types and correlation with expectations over others' behavior should thus mute the importance of self-image concerns in *visibility+bonus* compared with *visibility*. Using the theory to make more precise predictions for how *visibility+bonus* behavior will compare with behavior in the other treatments would require further assumptions about the relative magnitudes of the variance and covariance terms. In addition, to generate a proposition regarding the substitution effect from the theory, one would have to make stronger functional form assumptions about $E[win]_q$, $E[win]_w$, Ψ and Ω.

Conceptually, if the self-signal is valuable, then we expect crowding out when moving from *visibility* to *visibility+bonus*. This means we would see less socially desirable behavior when visibility is accompanied by a bonus, compared to visibility

alone. In the theory, this corresponds to $\gamma_q \varepsilon > \frac{(BE[win]_q)}{\varepsilon} + \gamma_q \Omega(q)$. If, in addition, q and $(w-q)$ are substitutes, we will also see increases in w^* with decreases in q^*, such that $\gamma_w \kappa < \frac{(BE[win]_w)}{\kappa} + \gamma_w \Psi(w)$. This leads us to the last two propositions.

Proposition 3 *In the* visibility+bonus *treatment, image concerns exceed marginal gains from getting a bonus from q, leading to crowding out. q^* will thus be greater in in the* visibility *treatment than it is in the* visibility+bonus *treatment.*

Proposition 4 *A drop in q^* between the* visibility *and* visibility+bonus *treatments will correspond to an increase in w^*.*

If (q^*, w^*) in the *visibility* and *visibility+bonus* treatments are not different, it means that the self-signal perfectly off-sets the potential gains from winning the bonus, especially through large increases in w.

4 Study design

4.1 Sample and field context

We investigate these ideas among judicial professionals in Tajikistan. Tajikistan is a lower middle income former Soviet country in Central Asia, with a GDP per capita of USD $2,834 (PPP, 2015 USD). According to the Polity IV index, Tajikistan is a relatively autocratic country. In the World Bank's *Doing Business* 2016 report, it

ranks 128th of 190 countries. It is a highly politicized environment, where delivering on quality may be a lower priority than being politically prudent.[5] Transparency International ranks it as 151st of 176 countries and territories for perceived levels of public sector corruption, in company with Uganda.[6] Likewise, the judicial system is also known for being corrupt (Bertelsmann Stiftung, 2014). The EBRD and World Bank's Business Environment and Enterprise Performance Survey (BEEPS) 2012 finds that only 35% of firms interviewed from a nationally representative sample agree with the statement that "the court system is fair, impartial and uncorrupted".

Tajikistan has a civil law legal system. It is comprised of the Constitutional Court, Supreme Court, Supreme Economic Court and a number of regional, city and district courts. The Constitutional Court decides on the constitutionality of legislation and presides over all other courts in the country. Next in the hierarchy are the Supreme Court, which oversees 3 regional courts and 68 city and district courts, and the Supreme Economic Court, which oversees 4 regional economic courts.[7] Prerequisites for becoming a judge include 5 years of legal education (this is equivalent to a 5 year undergraduate degree with a major in law and legal studies in the United States), 3 years of experience and being at least 25 years old. Students must also pass a qualification exam and successfully complete a one year internship. Judgeships are appointed by an examination board. Judicial appointments in Tajikistan are for 10 years, with the possibility of renewal. Appointees can expect to earn the average public sector salary (i.e. comparatively low) and receive additional fringe benefits such as a house and discounts on utilities (Stalbovskiy, Stalbovskaya and

[5] http://www.doingbusiness.org/rankings

[6] 2016 Corruption Perceptions Index, http://www.transparency.org/cpi2016

[7] There also exist 4 military or garrison courts that solely deal with cases related to the military.

Abdulhamidov, 2015).

Judicial decisions in Tajikistan are not publicly available, there are no independent legal publications in Tajik and "quarterly" newsletters from high courts are published only sporadically (European Bank for Reconstruction and Development, 2012). Accordingly, there is little opportunity for judges to establish a quality-based reputation. Further, extrinsic incentives for quality are poorly delineated. The courts operate under a tenure-based pay scale and once judges are appointed they face uncertainty with regards how performance is measured and how to secure reappointment (American Bar Association, 2008). The reappointment process is not governed by any law, lacks any objective measure, and thus lacks transparency. There is no system in place to track the number of decisions passed down by each judge or the quality. There is no statute of limitations for deciding a case once it has been filed, so judges cannot simply let cases die. As one might expect given these institutional details, volume performance in the country is low and Tajik speed of justice is poor (Colman, 2011).

Such uncertainty and low accountability can be highly corrosive to professionalism and quality. For example, the risk of not getting reappointed can lead to strategic decision making. In Tajikistan, this strategic decision making is apparent in any case involving the government, where the lack of independence in the judiciary makes the possibility of a foreign or domestic firm winning a case against the government nearly impossible (European Bank for Reconstruction and Development, 2012).

This context is not unique to Tajikistan and begs a review of the question "What

do judges maximize?" (Posner, 1993). In this setting, self-image may be key to motivating professional behaviour. Since there are so many factors that can undermine professionalism in the field – lack of visibility, job uncertainty, low fixed-rate salaries and capture by the executive branch – using the lab to identify whether or not quality may be valued within the profession is an important first step for understanding the potential power of increased visibility for improving decisions. While the presence of shirking, strategic decision making and corruption indicates a lack of professionalism, it does not preclude it. Some judges will perform well regardless of the circumstances. Moreover, professional norms concerning quality, even comparatively weak ones, are critical for the success of any program of using visibility to increase professionalism.

In the case of judges, quality can be universally defined as appropriate application of the legal code (effort), combined with an absence of rent-seeking. Effort cannot be directly incentivsed, and penalizing low quality work can have the unintended consequence of creating backlogs. Were decisions visible in this judicial system, quantity would be the more verifiable of the two dimensions, but since decisions are not visible, judges optimize over quantity and quality according to their own preferences for performance and professional norms.

Our sample for this study includes 103 commercial and general jurisdiction court judges, 57 from in and around Dushanbe, the capital city, and 46 from rural areas. At the time this constituted approximately 50% of the judges in the country that hear commercial cases. The average participant was a 39 year old male with 5 years of legal education and 12 years' experience in the legal sector. This is reported in

Table 3. We estimate that judges see around a dozen cases a month, including small procedural matters.

With this sample we contribute to a small body of literature on judge performance in former soviet countries. Dimitrova-Grajzl et al. (2012, 2015) investigate productivity among judges in Slovenia and Bulgaria using judge-level administrative data. While their results are only correlative, they find evidence of a quantity-quality trade-off in Slovenia, but not Bulgaria. They also find that volume is correlated with salary and the likelihood of promotion. These results echo our own findings – judges unambiguously respond to increased incentives by increasing volume. Unlike these papers, we are able to directly observe causal implications for quantity and quality of different incentives.

4.2 The experiment

The main treatment of interest in the experiment manipulates visibility of actions to activate image concerns. The literature shows that people behave differently when they think or know that they are being watched. An audience can increase generosity and cooperative behavior, even when the watcher is only suggested (Haley and Fessler, 2005). Psychologists suggest that this is due to the internally generated self-image rewards from behaving consistent with social norms (Gleitman, 1996; Izuma, Saito and Sadato, 2010). Results extend beyond generosity and cooperative behavior: providing an audience has been shown to decrease anti-social behavior (Ernest-Jones, Nettle and Bateson, 2011) and increase the likelihood of voting

(Panagopoulos, 2014).

We use lab experiments since the research question cannot be addressed using available field data. Since all judges faced the same tasks, the experiment provides information that we cannot observe in the real workplace setting, where all cases are different. The experiments took place in August 2013 in a classroom at the Judicial Training Center in Dushanbe, Tajikistan. The experiment consisted of a real-effort task, a multiple price list activity (hereafter referred to as MPL) to elicit preferences over ambiguity and a post-experiment survey. All activities were conducted using paper and pen. We reviewed informed consent and experiment instructions verbally and on paper before the start of the experiment. In all sessions, work was completely anonymous; judges were given a numeric ID card upon entering the classroom and used stickers to affix this ID to each worksheet they filled out. Performance on the real-effort task and the MPL exercise determined final earnings and judges were fully briefed on how payments would be determined before the experiment began.

The experiment was a two-by-two between subjects design, summarised in Table 2. Each session had one of four treatments: no peer review (*control*), positive unknown probability of peer review (*visibility*), and each of these interacted with a tournament bonus incentive (*bonus*, and *visibility+bonus*). For the peer review treatment, instructions included a statement that there was a chance that subjects' work would be viewed by peers at the end of the session, but that their work would remain anonymous. This statement was also written on a black-board and we called attention to this detail while reviewing instructions verbally. We also demonstrated how completed task forms would be chosen - by randomly drawing a number of forms

from the pile of completed forms. Neither subjects nor the experimenter knew the probability that any given form would be selected for review. We thus refer to the treatment as uncertain peer review rather than risky peer review. For treatments without the possibility of peer review, this detail was absent.

All review occurred at the end of the experiment. The review consisted of publicly randomly selecting some completed and graded worksheets and making them available for subjects to go over on their own while we made the subject payments. Bonus earnings were not marked on the worksheets and were private information of those who won the bonus.[8]

Thus *visibility* is a manipulation using what Kevin J. Haley and Daniel M.T. Fessler (2005) refer to as "subtle cues of observability"; we prime subjects to consider what others might think of their behaviour, while maintaining all subjects' anonymity. Since all tasks were completed before any review occurred, we measure the impact of the *possibility* of peer review, not the impact of the review itself. Our work goes further than the eye spots studies, however, in that our potential observer is an actual person. The potential observers, moreover, share a group identity with the subject.

For the bonus treatment, instructions included a modification of how earnings would be determined; judges learned that the top three point earners of the session would get a 50 point bonus. We chose a tournament bonus because we wanted to incentivize volume and did not have a prior on a meaningful threshold level to set

[8]Graded worksheets had pluses marked next to each correct answer and minuses marked next to each incorrect answer. The top of the graded worksheet listed points earned from correct items, points earned from incorrect items, and total points earned.

ex ante. Further, a competitive bonus rather than a threshold bonus requires subjects to consider the performance they expect from their peers, thus creating a parallel with the self-image incentive where subjects also may reflect on what their peers are doing. Note that the bonus was based on points earned (see below for a description of how points were earned) and not number of items completed correctly.

The real-effort task we used entailed filling in blanks from excerpts of the Tajik commercial code, where we had removed key words or phrases, or crossing out extraneous words or phrases that we had inserted. Items covered a range of difficulty. A copy of the Tajik commercial code and a recently used training manual were provided to everyone. Thus the task mimicked key aspects of the judges' real work environment: it used relevant subject matter, there was no required minimum or maximum effort and judges could use references as desired. The task was mundane and simple enough that anyone could succeed, but impossible to complete in the time provided.

Subjects had 8 minutes to complete as many items on the worksheet as they could. Worksheets were designed to be impossible to complete all items correctly but such that anyone could have completed at least one item correctly. Subjects earned 5 points for any attempt at filling in the blank, and an additional 5 points if they did so correctly. This simplifies to 10 points for each correct answer and 5 points for each incorrect answer. Subjects earned 1 Tajik somoni (0.15 euros) for every 10 points.[9] Note that unlike the "rational cheater" model of motivation, the wrong answers in this experiment are not considered cheating as quality is fully

[9]Participants completed an incentivized practice round so that they could familiarize themselves with the task. Data from the practice round are not included in the analysis.

observable, points are given for effort on both correct and incorrect answers and the agent's themselves are the only beneficiaries of the effort (or lack thereof).

After the real-effort task, while completed forms from the task were being scored, subjects completed the MPL activity and the survey. For the MPL activity, subjects made 11 choices. Each choice involved a trade-off between a safe payout and a payout that would be determined by a lottery. To determine the outcome for each lottery, we drew red and black balls from an urn. We capture ambiguity preferences by not specifying the proportion of red and black balls in the urn. Subjects were paid for each choice to minimize confusion. Unfortunately the subjects did not fully understand the MPL activity and we observe multiple switch points or no switch points for all but three subjects. Subjects did, however, understand that each choice was between a safe and uncertain lottery. Thus, we use the fraction of times the subject chose the lottery, instead of the safe option, as our measure of ambiguity preferences.

The post-experiment survey was not incentivized. It included items on demographics and asked subjects about strategies they may have used for completing the task. It also included a set of 12 items to measure grit (Duckworth et al., 2007).[10]

All payments were made at the end of the sessions. There was no feedback between activities, so participants did not learn how many points (i.e. how much money) they or their peers earned until the end of the session. Experiments were conducted in Tajik, with written instructions available in Russian.

[10]The grit score is the average response from a 12 item survey.

A final note on the experimental design: the visibility treatment includes a combination of anonymity and uncertainty. Subjects are told their peers present in the room may view their anonymised work. This anonymity is what ensures that we measure self image and not external reputation building. But visibility is uncertain, and the uncertainty is multifaceted: there is an unknown positive probability of information being posted and subjects do not know which of their peers may see any given decision. Different peers' opinions will have different value and peers have different abilities to verify the quality. Going back to the model, the best way to think of the visibility treatment is as an increase in x from zero to some non-zero value, down-weighted to reflect the uncertainty. To correct for the impact this may have on outcomes, we will control for ambiguity attitudes in the regressions. We also perform robustness checks where we interact ambiguity attitudes with treatment dummies. This yields no significant results and is available upon request.

4.3 Empirical specifications

To look at the impact of the treatments on effort choices, we first look at the treatment effects using Mann-Whitney tests, and complement this with OLS regressions. We compare the performance on total number of items completed, number of items completed correctly, and number of items completed incorrectly between treatments.

We also examine the question of how the different incentives impact the likelihood of a correct answer. This is a pertinent policy question. Users of the legal system rely on having a high probability of a just outcome, with reasoning behind

decisions properly based in the legal code.

Finally, we look at the trade-off between quantity and quality. In the multi-tasking literature, agents are hypothesized to put more (or all) effort into the easily measurable task. While we can perfectly observe both quality and quantity in our experiment, quantity is easier to verify and requires less effort to produce. We expect that the trade-off is more salient under the bonus incentive, where there are disproportionate gains to incorrect answers. The probability of a correct answer is also relevant when considering a quantity-quality trade-off because judges may have preferences over their average performance that does not change across treatments.

The regressions are as follows, where $Performance_i$ is either number correct, number incorrect, total items completed or the ratio of number correct to total items completed:[11]

$$Performance_i = \beta T_i + \gamma Z_i + \delta X_i + e_i. \tag{13}$$

Denoting total incorrect as h_i, the regression for the quantity-quality trade-off is:

$$h_i = \alpha(Numbercorrect_i) + \beta T_i + \gamma Z_i + \delta X_i + e_i. \tag{14}$$

T_i is a vector of treatment dummies, with the control as the excluded category. Z_i is a vector of judge characteristics including age, sex, education, seniority (years working in the legal sector), a dummy for being relatively ambiguity loving and a

[11] Other econometric specifications that account for non-normal error terms or count data yield the same results.

measure of grit.[12] X_i represents the day on which the subject participated to control for any information passed between subjects across days.

5 Results

Results are summarized in Table 4. First, it is important to review whether the incentives functioned properly, since pay-for-performance is not part of the judicial system in Tajikistan. While quality is poor in general, with half of the sample scoring below 85% correct in the control, there is considerable variation (individual values range from 0 to 100% correct). In all treatments we see subjects providing both correct and incorrect answers. The fact that we see correct answers, which have higher effort cost, in all treatments appropriately reflects the relative power of the incentives. Also, across treatments, subjects earned most of their points with correct answers. There were subjects who provided only correct answers in all treatments except for *bonus*, where all subjects provided a mixture of both correct and incorrect items. Thus, we conclude that despite the relative inexperience with experiments and with piece-rate incentives, subjects took the task and the incentive seriously.

We now move on to discuss the comparison of the treatments. Figure 1 displays the difference in average total, correct and incorrect items completed across treatments. Each bar represents the average number of items completed in each session.

[12]The total grit score has 12 indicators, but we found item non-response to be correlated with treatment. To avoid having non-random missing values for the score, we use one item from the questionnaire –"I am a hard worker" – as it directly captures the idea of perseverance, is strongly correlated with grit score (Pearson correlation of 0.66 and $p<0.001$) and non-response on this item was minimal and not correlated with treatment.

This is broken down by average number correct (bottom portion of the bar) and average number incorrect (top portion of the bar). Table 4 reiterates what is shown in Figure 1, and also includes additional statistics not pictured in the chart. Table 5 demonstrates the direction of the differences between treatments and their significance according to Mann-Whitney tests. Results are reiterated in the regression analysis, and also prove to be robust to adding subject and session characteristics.

Overall, we see that incentives work; quantities of items completed are higher in all the treatments compared with the control. We discuss more detailed results, corresponding to each proposition, in turn. First, we consider the impact of visibility. We see that the average number of correct items is higher in *visibility* and *visibility+bonus* than in *control*. *Visibility* and *visibility+bonus* also outperform *control* with respect to the average number of incorrect items. This gives our first result.

Result 1 (visibility improves effort) *Subjects increase both correct and incorrect items when their work is subject to anonymous visibility, compared to the control.*

This result provides evidence that both γ_q and γ_w are greater than zero. This means that judges receive image gains from increases in both quantity and quality. We cannot distinguish if increases in $(w-q)^*$ are deliberate or the natural consequence of trying for a higher q^* but with limited capacity. That being said, participants had the ability to look up the answers in order to ensure they get all attempted items correct. Moreover, we fail to reject the hypothesis that q^* is the same between the two peer review treatments (Mann-Whitney test, p=0.43). This underlines the idea that total productivity, w^*, matters for self-image and that, in this work setting,

there is some acceptable level of $(w-q)^*$.

Second, we consider the impact of a bonus tournament. In *bonus*, quantity completed goes up much more, compared with the control, than in *visibility* and *visibility+bonus*. This is due to a large increase in average number of items completed incorrectly in *bonus*. Aside from the striking differences in the total items completed, *bonus* motivated a large decrease in the percent of items completed correctly compared with the other three groups. What is perhaps remarkable is that q^* in *bonus* is not statistically different from q^* in *control*. This leads to our second result.

Result 2 (a competitive bonus increases rent-seeking) *In contrast to visibility, a competitive bonus incentive motivates an increase w^* with no difference in q^* compared to the control.*

This result is consistent with $E[win]_w > 0$, but is contrary to Proposition 2, which predicts increases in q. Thus, in this sample, we cannot reject the possibility that $E[win]_q = 0$. Even though there is an increase in rent seeking, subjects also invest some minimum effort in quality. Moreover, this behavior is sensitive to visibility. Combining visibility with a competitive bonus seems to fully offset the rent-seeking incentive of the bonus. This gives our next result.

Result 3 (visibility crowds in good behavior) *Subjects complete fewer incorrect items in the visibility+bonus treatment than they do in the bonus treatment. Combining visibility with a competitive bonus has the same impact on socially desirable behavior (number of correct items and per cent correct) as visibility alone has.*

Self-image concerns thus enforce a quality standard that would otherwise be ignored in the presence of a bonus. It bears mentioning that female judges and judges with more experience tend to have higher percentages correct. Looking back at Table 6 we can see that this is not due to answering more correctly, but instead is due to answering fewer incorrectly.

We now turn to investigating the quantity-quality trade-off across the treatments. This is partly informed by comparing performance on percent correct. Results appear in Table 5 and Table 7.[13] In *bonus* we see that subjects provided 1.6 incorrect answers to each correct answer, compared with a 0.44 ratio in *control*. The ratio is middling in *visibility* and *visibility+bonus* (0.78 and 0.60, respectively) and not significantly different than in *control*. Interestingly, the percent correct does not differ between *visibility* and *visibility+bonus*, nor does it differ when comparing these two treatments to *control*.

More interesting is the correlation between the number that subjects answer correctly and the number they answer incorrectly. The latter correlations are in Table 8, where we use interaction terms to look at how this correlation differs between treatments. We see that the number correct and the number incorrect are significantly negatively correlated on average. In all treatments people with more correct answers tend to also have fewer incorrect answers. This suggests a trade-off between quantity and quality. Finally, the trade-off is the same in the control and in the visibility treatments, but it is different in the bonus treatment. In the bonus treatment, subjects are more willing to gain more points from incorrect answers. We see from

[13] One subject choose to complete zero items and so their ratio of correct to total is undefined.

the results in columns 3 and 4 of Table 8 that in *bonus* this trade-off approaches 1. This means that more people are willing to have a higher incorrect-to-correct ratio in *bonus*. The range of ratio in *bonus* is 0.37 to 46. In the control it is 0 to 5. This leads to our final result.

Result 4 (correct and incorrect answers are substitutes) *Controlling for treatment, a decrease in the number of correct answers by 1 results in 0.3 additional incorrect answers. In the bonus treatment, there is a threefold increase in willingness to substitute foregone correct answers with incorrect answers, compared to the control.*

The trade-off described is not as we would expect from solving a straightforward profit maximization problem. If subjects were simply maximizing profit subject to a time constraint, and given the pay-offs, we would expect the coefficient on *Correct* in Table 8 to be equal to -2. This can be seen if we let profit be $10C + 5I$, where C is the number correct and I is the number incorrect. Let $T = T_C C + T_I I$ be the time constraint, where T is total time available, T_C is time spent on each correct answer and T_I is time spent on each incorrect answer. Given this decision problem, we would expect subjects to spend twice as much time on correct answers as incorrect answers (i.e. $1/2 = T_I T_C$). Solving the time constraint for I in terms of C then gives $I = T/T_C - T_C/T_I C$ or $I = T/T_C - 2C$.[14] But we do not see this in the data. Instead, we see from our regression of *Incorrect* on *Correct* that the coefficient is -0.3.

From this we conclude that the optimization problem subjects face in this exper-

[14]The same result obtains if we allow subjects to choose an effort level to achieve a certain probability of getting any given item correct, or if we solve for choosing quantity and quality, where both quantity and quality are linear combinations of C and I.

iment is likely more complex than basic profit maximization. Self-image utility can explain the apparent limit on willingness to earn points from incorrect answers. In a world where subjects have positive self-image utility from quality, for each additional item completed they decide the effort to put forth so as to achieve a quality that maximizes the self-image payout. Meanwhile, a negative self image value to appearing greedy leads subjects to replace each foregone correct with a disproportionately small number of incorrect items. Results hold even when we control for observables that will be correlated with ability to answer items correctly, such as education and years of experience. They also hold if we exclude outliers, who may be strongly driven by the competitive nature of the bonus and thus driving the result.

In summary, we find that image concerns do motivate more effort, but the quantity-quality trade-off does not appear to be affected by such rewards. The (private) bonus treatment, on the other hand, overemphasizes gains to incorrect answers that are easier to complete in a fixed time frame - here is the only place we see the quantity-quality trade-off suffer. The impact of the bonus on incorrect items completed is approximately four times the effect that visibility has on completing correct items. Percent correct is also higher in all treatments compared to the *bonus*. This means that the probability that a court case will be handled well, instead of being handled quickly with errors, predictably decreases under the competitive bonus incentives. Adding self-image to a bonus incentive scheme mitigates the opportunism we see in the private bonus treatment and is not discernible from self-image rewards alone. Finally, that visibility reduces the effect of the bonus suggests that image rewards can counteract extrinsic motivation along a socially desirable dimension.

6 Conclusion

This paper explores how self-image and monetary rewards motivate professionals to expend effort on quantity and quality. Understanding how self-image influences effort in the workplace is particularly important because these factors can compensate for the inability to contract over quality. But this has been sparsely explored among professional, who face different motivations and norms. We investigate this topic using a framed field experiment with sample of judges in Tajikistan. Creating incentives for quality and professionalism in a setting such as Tajikistan is difficult. Once judicial decisions have been passed down, they are not open for peers or the public to access. Further, income is tenure-based and promotion is not linked directly to performance. Our paper suggests that leveraging judges' professional self-image is an alternative policy option in such a setting.

Our results support the idea that self-image can influence performance and under certain conditions it requires a stimulus. We find that both monetary and self-image incentives increase effort put towards quantity. Under the self-image incentive, the additional effort does not come at the cost of quality. The bonus incentive, on the other hand, is very bad for quality. While increasing the strength of the monetary incentive (by adding a competitive bonus) does not decrease quality, it does increase undesirable behavior. So the bonus has adverse effects compared to the control. Extrapolating to the real world, such bonus incentives could actually decrease the chance that any given commercial law decision is of high quality, without a complementary social incentive. Also, from a user perspective, visibility alone may be an appropriate policy option when judges display limited preferences for quality:

the chances of a fair outcome are no worse and the volume of cases considered in a given time frame could be better, thus improving the speed of justice.

Self-image and bonus incentives together motivate quality better than a pure money market with a bonus. In fact, behavior in the mixed treatment is indistinguishable from behavior in a self-image reward treatment with no bonus. The extra money does not crowd out good behavior. Instead, adding visibility to a bonus crowds in good behavior compared to bonus alone. Note that our results from professionals in a work environment correspond in part with James Heyman and Dan Ariely (2004)'s work on charitable acts, which finds that mixed markets more closely resemble money markets. This is reflected in the lack of difference in percent correct in the control (pure money market) and the self-image rewards treatments (mixed market). But in contrast to the conclusion that money markets resemble mixed markets, we find that mixing the (socially-oriented) self-image incentive with the high-powered bonus more closely resembles the the visibility treatment (a mixed market) than the control, a pure money market. So while it is not advisable to implement the competitive bonus, per se, if considered it must be accompanied by additional incentives for professionalism.

These results should be carefully considered in any policy aiming to increase visibility. Visibility policies are usually not anonymous, so there are additional reputational incentives to take into account. Concerns over reputation may create enough social pressure to motivate improvements in average quality that we do not see in the anonymous setting. On one hand, reputation may make career concerns more salient, motivating competition over quality that is not present otherwise. On the

other hand, if the self-image response to anonymous visibility reflects professional norms, it is not clear that reputational incentives will change the quantity-quality trade-off that exists presently.

We note that increasing anonymous visibility is essentially a reduced form treatment. The message is sufficient to trigger self-image investment, but we cannot disentangle what elements of the message do this. With our model we assume that this is due to preferences over both quality and quantity, where quantity is emphasized because it is more easily verified by peers. But it may also be capturing career concerns, if the message additionally cues the belief that not only peers, but also authorities will see the work.

Our results best generalize to the work place for the marginal worker, who prefers to conform to quality norms but is too often is uninspired to do so. In the real work place, increasing volume may require a different kind of effort and gains from rent-seeking may be a relatively large portion of income. But according to the theory, the social stimulus is a motivator that should work across contexts. Undoubtedly some judges persistently perform better than others, especially those for whom the professional identity is more often salient. For others, monetary opportunity costs may act as a constraint on norm conformance – the preference may be to perform according to professional norms, except when costs are high as judges may determine it is not worth the effort. But for the marginal worker, reminders of professional norms may hold the key to motivating more professional behavior.

Table 1: Participant characteristics

	Median	Mean	St. Dev.	Min	Max	N
Age	35	39.25	10.18	26	64	103
Sex (female=1)	0 (Male)	0.18	0.39	-	-	98
Years of legal education	5	5.24	0.88	3	9	102
Years of experience	10	12.28	7.87	2	36	98

Table 2: Two-by-two design

	No chance of visibility	Chance of visibility
No bonus	*control*	*visibility*
Bonus	*bonus*	*visibility+bonus*

Table 3: Participant characteristics, by treatment

	N	Age	Sex (Female=1)	Years of legal education	Years experience in legal sector
Control	25	38.32 (9.87)	0.21 (0.41)	5.08 (0.76)	10.65 (5.44)
Visibility	23	37.96 (9.42)	0.18 (0.39)	5.09 (0.95)	13.35 (7.72)
Bonus	15	38.20 (10.37)	0.29 (0.47)	5.20 (0.41)	12.40 (7.24)
Visibility + bonus	40	40.98 (10.84)	0.13 (0.34)	5.44 (1.02)	12.69 (9.40)
p-value from largest pairwise difference		0.270	0.229	0.140	0.182

Note: Standard deviation is reported in parenthesis. The last row reports the p-value from pairwise two-sided t-tests of differences in means between treatments. For the sex variable, we used Fisher's exact test. Four judges with 8 and 9 years of education in the *visibility+bonus* treatment drive the imbalance in years of legal education across treatments. Results do not change if we control for education or drop them from the sample.

Figure 1: Average performance across treatments

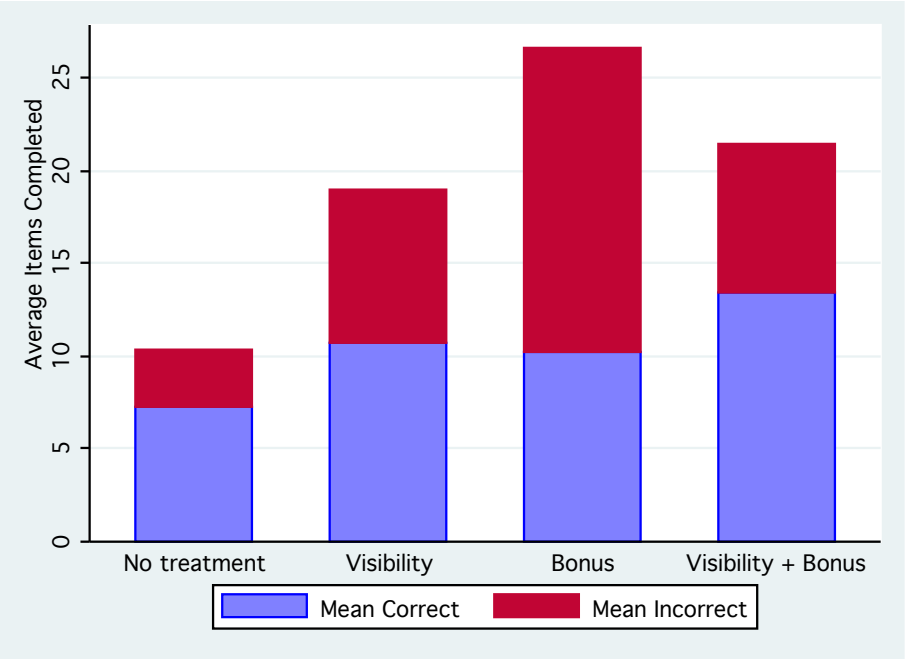

Table 4: Results summary, by treatment

	Mean items completed	Mean correct	Mean incorrect	% correct	% with C > I
	w	q	$w-q$	q/w	
Control	10.36	7.2	3.16	70.21%	72%
	(4.42)	(3.96)	(3.50)	(28.32)	(45.83)
Visibility	19	10.70	8.30	60.14%	65.21%
	(11.15)	(5.76)	(10.04)	(28.47)	(48.70)
Bonus	26.6	10.2	16.4	42.54%	47%
	(9.96)	(6.42)	(12.66)	(26.06)	(51.64)
Visibility + bonus	21.4	13.35	8.05	62.12%	73%
	(11.25)	(10.25)	(8.42)	(27.96)	(45.22)

Note: Standard deviations are reported in parentheses.

Table 5: Mann-Whitney results:
+ means column is more often higher than row,
- means that row is more often higher than column

		Visibility	Bonus	Visibility + bonus
Total items completed, w	Control	+ (0.001)	+ (0.000)	+ (0.000)
	Visibility		+ (0.009)	+ (0.331)
	Bonus			- (0.080)
Correct items, q	Control	+ (0.025)	+ (0.161)	+ (0.009)
	Visibility		- (0.730)	+ (0.483)
	Bonus			+ (0.426)
Incorrect items, $w-q$	Control	+ (0.007)	+ (0.000)	+ (0.001)
	Visibility		+ (0.002)	- (0.688)
	Bonus			- (0.002)
Percent correct, q/w	Control	- (0.109)	- (0.002)	- (0.119)
	Visibility		- (0.042)	- (0.887)
	Bonus			+ (0.012)

Note: p values from Mann-Whitney rank sum tests are reported in parentheses. The hypotheses tested are that the ranks of total items completed, correct items, incorrect items and percent completed correctly are, respectively, equal across treatments.

Table 6: Regression results, ordinary least squares

	(1) Total completed, w	(2) Incorrect, $w-q$	(3) Correct, q	(4) Total completed, w	(5) Incorrect, $w-q$	(6) Correct, q
Visibility	8.640	5.144	3.496	8.521	5.048	3.473
	(2.495)	(2.213)	(1.442)	(2.761)	(2.383)	(1.452)
Bonus	14.133	12.516	1.616	14.987	14.037	0.950
	(3.198)	(3.621)	(2.374)	(3.632)	(4.048)	(2.579)
Visibility + bonus	11.356	4.999	6.358	10.669	3.379	7.290
	(1.933)	(1.622)	(1.713)	(2.657)	(1.722)	(2.337)
Age				0.137	0.377	-0.241
				(0.170)	(0.151)	(0.126)
Female				-3.539	-7.806	4.267
				(3.933)	(2.053)	(3.182)
Education				1.124	0.339	0.784
				(1.004)	(0.815)	(0.786)
Experience				-0.289	-0.468	0.180
				(0.184)	(0.168)	(0.126)
Ambiguity attitude				-4.434	-3.272	-1.162
				(3.215)	(2.830)	(2.330)
Work ethic				-0.217	-0.153	-0.064
				(1.082)	(0.835)	(0.777)
Constant	6.145	1.712	4.433	4.186	-3.352	7.538
	(3.546)	(3.007)	(3.136)	(10.917)	(6.889)	(9.312)
r2	0.251	0.182	0.104	0.304	0.322	0.199
N	103	103	103	92	92	92

Note: Robust standard errors are in parentheses. Regressions also include a control for the day the session occurred.

Table 7: Quantity-quality tradeoff, Percent correct, (q/w), as the dependent variable

	(1)	(2)
Visibility	-0.101	-0.062
	(0.082)	(0.084)
Bonus	-0.324	-0.368
	(0.098)	(0.106)
Visibility + bonus	-0.074	-0.035
	(0.072)	(0.075)
Day of session	0.047	0.063
	(0.045)	(0.043)
Age		-0.012
		(0.004)
Female		0.267
		(0.066)
Education		0.037
		(0.032)
Experience		0.013
		(0.004)
Ambig. attitude		0.097
		(0.083)
Work ethic		0.011
		(0.031)
Constant	0.608	0.266
	(0.106)	(0.284)
R-square	0.097	0.290
N	103	92

Note: Robust standard errors are in parentheses.

Table 8: Quantity-quality tradeoff: Number Incorrect, $(w - q)$, as the dependent variable

	(1)	(2)	(3)	(4)
Correct	-0.337	-0.288	-0.265	-0.043
	(0.127)	(0.146)	(0.127)	(0.261)
Visibility	6.322	6.047	4.772	4.437
	(2.295)	(2.523)	(3.853)	(5.162)
Bonus	13.061	14.310	22.891	25.486
	(3.330)	(3.786)	(6.200)	(6.590)
Visibility + bonus	7.140	5.475	6.461	5.584
	(1.983)	(1.929)	(3.286)	(3.617)
Correct x Vis.			0.122	0.057
			(0.244)	(0.420)
Correct x Bonus			-0.973	-1.182
			(0.402)	(0.510)
Correct x (Vis. + bon.)			0.017	-0.140
			(0.189)	(0.317)
Constant	3.205	-1.185	2.935	-5.487
	(2.682)	(6.815)	(3.289)	(6.843)
Controls	No	Yes	No	Yes
R-square	0.252	0.366	0.310	0.428
N	103	92	103	92

Note: Robust standard errors are in parentheses. The controls included in columns (2) and (4) comprise *age*, *gender*, *education*, *years of experience in the legal sector*, day of the sessions, as well as measures of ambiguity preferences and work ethic (from the Grit questionnaire).

References

Akerlof, George A, and Rachel E Kranton. 2000. "Economics and identity." *Quarterly Journal of Economics*, 715–753.

Akerlof, George A., and Rachel E. Kranton. 2008. "Identity, Supervision, and Work Groups." *The American Economic Review*, 98(2): pp. 212–217.

American Bar Association. 2008. *Judicial Reform Index for Tajikistan.* Washington, DC:American Bar Association.

Andreoni, James, and B Douglas Bernheim. 2009. "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects." *Econometrica*, 77(5): 1607–1636.

Ariely, Dan, Anat Bracha, and Stephan Meier. 2009. "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review*, 544–555.

Ashraf, Nava, Oriana Bandiera, and B Kelsey Jack. 2014. "No margin, no mission? A field experiment on incentives for public service delivery." *Journal of Public Economics*, 120: 1–17.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2013. "Team incentives: evidence from a firm level experiment." *Journal of the European Economic Association*, 11(5): 1079–1114.

Bem, Daryl J. 1972. "Self-perception theory." *Advances in Experimental Social Psychology*, 6: 1–62.

Bénabou, Roland, and Jean Tirole. 2003. "Self-knowledge and self-regulation: An economic approach." *The Psychology of Economic Decisions*, 1: 137–167.

Bénabou, Roland, and Jean Tirole. 2006. "Incentives and prosocial behavior." *The American economic review*, 96(5): 1652–1678.

Bertelsmann Stiftung. 2014. *BTI 2014 - Tajikistan Country Report.* Gütersloh:Bertelsmann Stiftung.

Bodner, Ronit, and Drazen Prelec. 2003. "Self-signaling and diagnostic utility in everyday decision making." *The Psychology of Economic Decisions*, 1: 105–26.

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic incentives and social preferences: substitutes or complements?" *Journal of Economic Literature*, 368–425.

Carpenter, Jeffrey, and Caitlin Knowles Myers. 2010. "Why volunteer? Evidence on the role of altruism, image, and incentives." *Journal of Public Economics*, 94(11): 911–920.

Charness, Gary, David Masclet, and Marie Claire Villeval. 2013. "The dark side of competition for status." *Management Science*, 60(1): 38–55.

Cohn, Alain, Ernst Fehr, and Michel André Maréchal. 2014. "Business culture and dishonesty in the banking industry." *Nature*, 516: 86–89.

Colman, Alan. 2011. "Court decisions in commercial matters: an EBRD assessment." *Business Law International*, 12(2): 155–184.

DellaVigna, Stefano, and Devin Pope. 2016. "What motivates effort? Evidence and expert forecasts." National Bureau of Economic Research.

Dimitrova-Grajzl, Valentina, Peter Grajzl, Katarina Zajc, and Janez Sustersic. 2012. "Judicial incentives and performance at lower courts: evidence from Slovenian judge-level data." *Review of Law & Economics*, 8(1): 215–252.

Dimitrova-Grajzl, Valentina P, Peter Grajzl, Atanas Slavov, and Katarina Zajc. 2015. "Courts in a Transition Economy: Case Disposition and the Quantity-Quality Tradeoff in Bulgaria." Available at SSRN 2569351.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: perseverance and passion for long-term goals." *Journal of Personality and Social Psychology*, 92(6): 1087–1101.

Ernest-Jones, Max, Daniel Nettle, and Melissa Bateson. 2011. "Effects of eye images on everyday cooperative behavior: a field experiment." *Evolution and Human Behavior*, 32(3): 172–178.

European Bank for Reconstruction and Development. 2012. *Commercial Laws of Tajikistan, April 2012, An Assessment by the EBRD.*

Frey, Bruno S, and Felix Oberholzer-Gee. 1997. "The cost of price incentives: An empirical analysis of motivation crowding-out." *The American Economic Review*, 746–755.

Gleitman, Henry. 1996. *Basic Psychology.* New York:Norton.

Gneezy, Uri, and Aldo Rustichini. 2000. "A Fine is a Price." *The Journal of Legal Studies*, 29(1): 1–17.

Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel. 2011. "When and why incentives (don't) work to modify behavior." *The Journal of Economic Perspectives*, 191–209.

Haley, Kevin J., and Daniel M.T. Fessler. 2005. "Nobody's watching?: Subtle cues affect generosity in an anonymous economic game." *Evolution and Human Behavior*, 26(3): 245–256.

Heyman, James, and Dan Ariely. 2004. "Effort for payment a tale of two markets." *Psychological Science*, 15(11): 787–793.

Hoff, Karla, and Priyanka Pandey. 2014. "Making up people - The effect of identity on performance in a modernizing society." *Journal of Development Economics*, 106: 118 – 131.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design." *Journal of Law, Economics, & Organization*, 24–52.

Izuma, Keise, Daisuke N Saito, and Norihiro Sadato. 2010. "Processing of the incentive for social approval in the ventral striatum during charitable donation." *Journal of Cognitive Neuroscience*, 22(4): 621–631.

Kosfeld, Michael, and Susanne Neckermann. 2011. "Getting more work for nothing? Symbolic awards and worker performance." *American Economic Journal: Microeconomics*, 3(3): 86–99.

Lacetera, Nicola, and Mario Macis. 2010. "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme." *Journal of Economic Behavior & Organization*, 76(2): 225–237.

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of political Economy*, 89(5): 841–864.

Nagin, Daniel S., James B. Rebitzer, Seth Sanders, and Lowell J. Taylor. 2002. "Monitoring, Motivation, and Management: The Determinants of Opportunistic Behavior in a Field Experiment." *The American Economic Review*, 92(4): pp. 850–873.

New, Peter Kong-ming, and Richard M Titmuss. 1972. "The Gift Relationship: From Human Blood to Social Policy."

Panagopoulos, Costas. 2014. "Watchful eyes: implicit observability cues and voting." *Evolution and Human Behavior*, 35(4): 279–284.

Posner, Richard A. 1993. "What Do Judges and Justices Maximize? (The Same Thing Everybody Else Does)." *Supreme Court Economic Review*, 3: 1–41.

Ryan, Richard M, and Edward L Deci. 2000. "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." *American Psychologist*, 55(1): 68.

Seabright, Paul B. 2009. "Continuous preferences and discontinuous choices: How altruists respond to incentives." *The BE Journal of Theoretical Economics*, 9(1).

Stalbovskiy, Oleg, Maria Stalbovskaya, and Bakhtiyor Abdulhamidov. 2015. "UPDATE: Law of the Republic of Tajikistan - A Guide to Web Based Resources." [Online; accessed 5-April-2017].

Turner, John C, and Rina S Onorato. 1999. "Social identity, personality, and the self-concept: A self-categorization perspective." *The Psychology of the Social Self*, 11–46.

A Mathematical Appendix

In this paper we specify a utility function, $U(w,q) = Y(w,q) + R(w,q) - C(w,q)$, where a worker choses effort to put toward quantity and quality, (w,q). $Y(w,q)$ is income from the effort choice, $R(w,q)$ is self-image gains (or losses) from the effort and $C(w,q)$ is the cost of the effort. This appendix provides more details on the results presented in the main text of the paper. First, we recall the functional forms for R, Y and C, and the terms in each function. Second, we show the first order conditionals and how these translate to treatment effects.

Self-image, $R(w,q)$, is defined by intrinsic valuations for quantity and quality, v_w and v_q, as well as by returns to appearing to care about quantity and/or quality, $E(v_q|q,g)$ and $E(v_w|w,p)$, respectively. x denotes the visibility (or salience) of the effort choice. It enhances the utility gains or losses from appearing to care about q and/or w.

$$R(w,q) = x\left[\gamma_w E(v_w|w,p) + \gamma_q E(v_q|q,g)\right] \tag{15}$$

with $\gamma_q \geq 0$. $\mathbf{v} = (v_w, v_q) \in \mathbb{R}^2$, where (v_w, v_q) is drawn independently from a continuous distribution with known mean and variance.

$$\begin{pmatrix} v_w \\ v_q \end{pmatrix} \sim N\left(\begin{matrix} \bar{v}_w \\ \bar{v}_q \end{matrix}, \begin{bmatrix} \sigma_w^2 & \sigma_{wq}^2 \\ \sigma_q^2 & \sigma_{wq}^2 \end{bmatrix} \right)$$

Income is a linear combination of quantity and quality, weighted by intrinsic valuation of each kind of effort. It is also a function of the expected value of winning a bonus, $E[win]$.

$$Y(w,q) = v_w pw + v_q gq + E[win] \tag{16}$$

where p is the per unit wage. g is a premium the worker earns for each item completed correctly. B is a bonus for being a top performer, where performance is the total value of the tasks completed, $pw+gq$. $E[win] = B*p[(pw+gq)_i \geq (pw+gq)_-i]$.

Lastly, cost is $C(w,q) = \kappa w^2/2 + \varepsilon q^2/2$.

The long form of the utility function is thus:

$$U(w,q) = v_w pw + v_q gq + E[win] + x\left[\gamma_w E(v_w|w,p) + \gamma_q E(v_q|q,g)\right] - C(w,q) = \kappa w^2/2 + \varepsilon q^2/2. \tag{17}$$

We let q be the numeraire and all formulas will be expressed in terms of p. The first order conditions for this model thus give:

$$w* = \frac{1}{\kappa}\left(v_w p + \frac{\partial E[win]}{\partial w} + x\gamma_w \frac{\partial E(v_w|w,p)}{\partial w}\right) \tag{18}$$

$$q* = \frac{1}{\varepsilon}\left(v_q + \frac{\partial E[win]}{\partial q} + x\gamma_q \frac{\partial E(v_q|q,g)}{\partial q}\right) \tag{19}$$

Using Bayesian updating results for Normal random variables (i.e. Kalman filter), we obtain

$$E[v_w|w,p] = \overline{v}_w + \frac{p^2\sigma_w^2 + Bcov(v_w p, E[win]_w)}{p^2\sigma_w^2 + B^2 var(E[win]_w) + 2Bpcov(v_w, E[win]_w)}(w - E(w))\kappa \tag{20}$$

and

$$E[v_q|q,g] = \overline{v}_q + \frac{\sigma_q^2 + Bcov(v_q, E[win]_q)}{\sigma_q^2 + B^2 var(E[win]_q) + 2Bcov(v_q, E[win]_q)}(q - E(q))\varepsilon \tag{21}$$

where $E[win]_w = \frac{\partial E[win]}{\partial w}$ and $E[win]_q = \frac{\partial E[win]}{\partial w}$, to simplify the notation.

Treatments are then defined by whether x and B equal 1 or 0. Derivations of $(w*, q*)$

for each treatment are shown below. Interpretations of these formulas appear in the main text of the paper.

Control (x=0, B=0):

$$w* = \frac{v_w}{\kappa} \tag{22}$$

$$q* = \frac{v_q}{\varepsilon} \tag{23}$$

T1, visibility (x=1, B=0):

$$w* = \frac{v_w}{\kappa} + \gamma_w \frac{\partial E[v_w|w,p]}{\partial w} \tag{24}$$

$$= \frac{v_w}{\kappa} + \gamma_w \kappa \tag{25}$$

$$q* = \frac{v_q}{\varepsilon} + \gamma_q \frac{\partial E[v_q|q,g]}{\partial q} \tag{26}$$

$$= \frac{v_q}{\varepsilon} + \gamma_q \varepsilon \tag{27}$$

T2, bonus (x=0, B>0):

$$w* = \frac{1}{\kappa}\left(v_w + B\frac{\partial E[win]}{\partial w}\right) \tag{28}$$

$$= \frac{1}{\kappa}\left(v_w + BE[win]_w\right) \tag{29}$$

$$q* = \frac{1}{\varepsilon}\left(v_q + B\frac{\partial E[win]}{\partial q}\right) \tag{30}$$

$$= \frac{1}{\varepsilon}\left(v_q + BE[win]_q\right) \tag{31}$$

T3, visibility + bonus (x=1, B>0):

$$w* = \frac{1}{\kappa}\left(v_w + B\frac{\partial E[win]}{\partial w} + \gamma_w \frac{\partial E[v_w|w,p]}{\partial w}\right) \tag{32}$$

$$= \frac{1}{\kappa}\left(v_w + BE[win]_w\right) + \gamma_w \Psi(w) \tag{33}$$

$$q* = \frac{1}{\varepsilon}\left(v_q + B\frac{\partial E[win]}{\partial q} + \gamma_q \frac{\partial E[v_q|q,g]}{\partial q}\right) \tag{34}$$

$$= \frac{1}{\varepsilon}\left(v_q + BE[win]_q\right) + \gamma_q \Omega(q) \tag{35}$$

where

$$\Psi(w) = \left(\frac{p^2\sigma_w^2 + Bcov(v_w, E[win]_w)}{p^2\sigma_w^2 + B^2 var(E[win]_w) + 2Bpcov(v_w, E[win]_w)}\right)\kappa \tag{36}$$

and

$$\Omega(q) = \left(\frac{\sigma_q^2 + Bcov(v_q, E[win]_q)}{\sigma_q^2 + B^2 var(E[win]_q) + 2Bcov(v_q, E[win]_q)}\right)\varepsilon \tag{37}$$

Lastly, since the incentives used in the experiment are tied to total correct and incorrect answers (and not percent correct), the theory uses w and q as choice variables. Solving for percent correct given $Y(w,q)$ yields non-linear expressions. We can still gain insight from the theory, however, by considering w and q as choice variables, with changes in the percent correct coming as a result of that optimization. That is, we can derive expressions for q^*/w^*. Predictions from such an exercise are consistent with results. Calculations are available upon request.

B Experiment Protocol and Instructions

B.1 Procedures for the experiment, August 2013

Before the session begins we set the desks with reference material, a pen, a set of instructions and a consent form. Desks are arranged to give space between participants. Worksheets each have items in a random order, so even if judges try to cheat they cannot get answers by looking at the desks near them. We do not use partitions between participants because our local collaborator thought it would trigger suspicion and distrust of the activity. Also before each session, we put ID cards in a random order, according to a list we generated electronically. This was to ensure that there was no relationship between type of participant and high or low ID numbers. For example, People who arrived first or who wanted to be first in line would be no more likely to get the ID "1" than those who arrived late or preferred to enter the room last. For the sessions that included visibility, we also wrote on a chalkboard at the front of the room a statement about visibility taken from the instructions (i.e. that there was a chance that subjects' work would be viewed by peers at the end of the session, but that their work would remain anonymous).

As judges came into the room, they signed in on a sign-in sheet. Next, a moderator handed each participant an ID card. Each ID card had a number on it and a sticker sheet stapled to it. The sticker sheet had 5 stickers, each with the ID on it, which would later be used to indicate their ID on the various worksheets they filled out during the session. The card also had a note on it telling them to keep the card in order to get paid: "This is your randomly assigned ID for this session today. Please keep your ID confidential. Do not lose this card - you will need it to collect your payment at the end of the session." Participants took a seat at any desk where there were materials set out. We asked them to not look at the materials until everyone had arrived.

Once all participants had arrived, we welcomed them, reading a welcome script that was included the general instructions. We went over the general instructions, which included information on the ID cards and what the stickers were for. Sessions with visibility included a slightly augmented set of general instructions, where we informed them that someone might see the work they did during the session.

The general instructions also included information on the consent form. We read the consent form out loud and answered any questions. Consenting subjects signed

the form and we collected them. We waited for everyone to be done with this process before proceeding.

Round 1: Practice round

We passed out specific practice round instructions, and read them out loud. Subjects were not informed that it was a practice round. It was important to ensure subjects took the task seriously from the start of the session so we chose to pay them for all work completed during the session. They had never done anything like it before and we took precautions to avoid disengagement. Paying for all tasks also helped avoid confusion.

We then passed out the paper worksheets that had the "fill-in-the-blank" task. We asked for each participant to show their ID car in order to obtain a worksheet. We also made sure that each participant put the sticker with their ID number on the worksheet at that time. Subjects kept their sheets face down until instructed to begin. We then reminded participants that we would give them 8 minutes to complete the task, and that we would give them time updates when there was 5, 2 and 1 minute left to work. We then asked them to turn over the worksheets and begin the task. They worked on the task in the time given. After 8 minutes, we announced that time was up and collected the worksheets.

Round 2: Treatment round

We passed out specific Round 2 instructions, and read them out loud. We then passed out the paper worksheets that had the "fill-in-the-blank" task. Each person received a new worksheet with a different set of items on it. We asked for each participant to show their ID car in order to obtain a worksheet. We also made sure that each participant put the sticker with their ID number on the worksheet at that time. Subjects kept their sheets face down until instructed to begin. We then reminded participants that we would give them 8 minutes to complete the task, and that we would give them time updates when there was 5, 2 and 1 minute left to work. We then asked them to turn over the worksheets and begin the task. They worked on the task in the time given. After 8 minutes, we announced that time was up and collected the worksheets and the reference material.

Scoring and marking the worksheets

We corrected people's work by hand while the participants did rounds 3a and 3b. Participant answers were compared to an answer key. Marks made on worksheets

were the same for all worksheets. Rules for marking were: put an x by incorrect items, put a check by correct items; at the top of the worksheet, write the number of correct items and the number of incorrect items; show the multiplication of the points with the work done, and the total points; mark with a red pen, so it is easily seen.

Round 3a: Multiple price list activity

We distributed instructions for the MPL activity, then read instructions out loud. The instructions contained an example. We did a real draw from an urn to illustrate the example. We used a small cardboard box to represent the urn. It was filled with red and black ping pong balls. We made sure that everyone got a look at the urn and the balls inside.

After we finished going over the instructions, we distributed decision sheets. We asked for each participant to show their ID car in order to obtain a decision sheet. We also made sure that each participant put the sticker with their ID number on the decision sheet at that time. Subjects kept their sheets face down until instructed to begin. We told participants that they would have 15 minutes to complete both the decision task and the debriefing survey. We told them we would give them time updates when there was 5, 2 and 1 minute left to work. We then asked them to turn over the decision sheets and begin the task. They worked on the task in the time given. When someone finished filling out the decision sheet, we gave them a debriefing survey; this allowed some people who needed more time for the decision task to take longer.

Round 3b: Debriefing survey

We passed out debriefing surveys as participants finished their decision task. Participants put an ID sticker on the survey and began immediately. After 15 minutes, we announced that time was up and collected the surveys. As participants completed the debriefing survey, we entered their MPL decisions into a pre-formatted worksheet.

Drawing the lottery outcomes

Once all survey answer sheets were collected, we announced that we would draw the lottery. One ball was drawn for each choice. We asked a different participant to draw the ball for each choice. This was so that the draw was completely transparent and participants could trust the outcome. As each ball was drawn we made a list of

the outcomes on the chalkboard. This was then entered into a pre-formatted Excel workbook, such that participants' lottery payouts were automatically calculated.

Discussion

After the lottery, participants had the chance to ask questions about the task and get feedback on their work from the legal expert on the research team. For the *visibility* and *visibility+bonus* treatments, the moderator randomly selected 2/3 of the completed worksheets that would be available for anyone to review during this time (as well as during the payment process). To make this transparent, the moderator selected worksheets publicly, without looking at the IDs or the point totals on the graded sheets. The original protocol had a more elaborate approach to the random selection of sheets to be available for review, but the field context and the time limits of the sessions made it impossible to do anything more systematic. Selected sheets were placed on a desk at the front of the room, set off to the side so that those wanting to look at them would not be uncomfortable.

Participants were given time to review worksheets and/or ask questions. Meanwhile one of the moderators prepared to payments. At the end of the discussion, and when payments were ready, we read the final instructions text, which included how to collect payment. We also made sure all worksheets were collected.

Distribution of earnings

Cash envelopes were prepared by the moderators based on the Excel table. This was so that payment could be double checked at the time was was given, by the experimenter, without referring back to completed worksheets. This made it easier to maintain anonymity of the worksheets and privacy of the payments, and to ensure that completed task forms remained anonymous.

In order to collect the correct payment at the end, subjects had to present their numeric ID to the experimenter. The experimenter could then match the ID with the points earned, which had been entered into the Excel file. Once a subject was paid, we made a mark on their ID card, so that it could not be used again.

Subjects were paid on-by-one and earnings were handed over in envelopes. Envelopes were opaque so that participants could not see through the paper to determine which bills were inside. Total earnings were written on the inside flap of the envelope so that the participant could see their earnings and also count the cash, if they wanted. There was also a receipt, which both parties signed. The receipt listed

their point and somoni total. Participants signed receipts to acknowledge payment. Participants left the room once they had been paid.

B.2 Experiment Instructions

General Rules *[all groups]*

Welcome and thank you for being here today. This is an interactive session for the purpose of research. We are studying the recent training in commercial law and also studying judicial decision making. You will receive a small per diem for being here today, since this is not part of the training and your participation is voluntary. The per diem amount is 5 somoni (TJS). If you follow the instructions for this interactive session carefully and make good choices you can earn additional money beyond the per diem, potentially as high as 80 somoni (TJS). You will be paid in private and in cash at the end of the session.

It is important that you do not talk, or in any way try to communicate, with other people during the session. If you have a question during the session, raise your hand and a moderator will come over to where you are sitting and answer your question in private.

This interactive session will consist of three independent/unrelated rounds. In the first two rounds, your task will be to complete a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items.

In each round, you will earn points for how you complete items in the round. At the end of this session, you will get paid in cash according to the number of points that you earn across all rounds. You will earn 5 points for any item you attempt, even if it is not completed correctly. You will receive an additional 5 points if your answer is correct. Finally, in some rounds there will be an additional opportunity to earn points, also related to how you do the tasks. This will be described in detail at the start of the round.

You may use the materials in front of you to assist you in completing the items. The manuals are from a recent EBRD sponsored training in Tajik commercial law and contain text from and interpretation of the law in the areas of property, credit and real estate. You will also see excerpts from the civil code pertaining to these same areas of law. Please note that checking the manual will help you get items correct, but will also take time away from the number of items you can attempt.

We will review the answers for all items at the end of the session. We will not review answers between rounds. When we do go over answers, all participants will gather in the same room and our legal expert will lead the discussion. This will be the opportunity for you to ask any questions about the material contained in the tasks.

The third round will consist of an interactive module and a short survey. Unlike in other rounds, you will not be **not** required to fill blank spaces, or deal with any commercial law. For the interactive module, you will earn money according to the decisions you make. We will explain the details of this module at the start of Round 3.

Each participant has been assigned a randomly selected ID number. Please keep this number confidential.

How earnings are determined

In addition to the per diem, what you earn today will depend on the points you collect from completing the tasks. You will receive 1 somoni (TJS) in cash at the end of the session for every **10 *points*** you have earned. For example, those that answer all questions in the first two rounds, and get all of them correct, thus will get the largest sum 84 somoni (TJS), plus earnings from Round 3 and the per diem. Those that answer all questions, but will get some of them incorrect, will get less. Finally, those that will just participate and will not answer any question will only receive only earnings from Round 3 and per-diem. However if these people will answer even a single item in the first 2 rounds, and even if they will get this wrong, they will still be paid additional sum for their effort.

All earnings will be paid out at the end of the session today. You will be asked to sign a receipt of payment, acknowledging that the earnings were from your participation in this interactive session only.

If you have questions at this time, please raise your hand and we will answer it for everyone to hear before the session begins. If you have any questions once the session has started, please quietly raise your hand and one of the moderators will come to you to answer your question. It is important that you do not talk with any of the other participants.

[Additional Instructions for Groups 2 and 3 ONLY]

The tasks you complete in this session may be viewed by another participant that is here today. This will occur when we go over the answers to the tasks together at the end of the session. Note that while everyone has the chance to have their work viewed by another participant, it is likely that some people's work will not be viewed. You do not know how many people's work will be viewed by someone else, or the portion likely to have their work viewed. This is true for all participants.

Round 1 *[practice round]*

We are passing out a worksheet. Please do not look at it yet. In this round you will complete the items on the sheet in front of you. You will earn 5 points for each item you attempt, and an additional 5 points for each item you complete correctly. So, correctly completed items are worth a total of 10 points. Incorrectly completed items are worth 5 points.

The worksheet has a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items. Each blank is one item, and each cross out is one item. They are all worth 5 points if you attempt them, and an additional 5 points if correct.

You will have 8 minutes for this round. We do not expect you to finish all the items on the worksheet. We will collect the worksheets at the end of the round to calculate the points. Please turn over the worksheet and begin.

Round 2 *[treatment round]*

[Control]
We are again passing out a worksheet. Please do not look at it yet. In this round you will complete the items on the sheet in front of you, as in the previous round, but the items on this worksheet are different. As before, you will earn 5 points for each item you attempt, and an additional 5 points for each item you complete correctly.

As before, the worksheet has a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items. As before, each blank is one item, and each cross out is one item. They are all worth 5 points if you attempt them, and an additional 5 points if correct.

You will have 8 minutes for this round. We do not expect you to finish all the

items on the worksheet. We will collect the worksheets at the end of the round to calculate the points. Please turn over the worksheet and begin.

[Bonus treatment]
We are again passing out a worksheet. Please do not look at it yet. In this round you will complete the items on the sheet in front of you, as in the previous round, but the items on this worksheet are different. As before, you will earn 5 points for each item you attempt, and an additional 5 points for each item you complete correctly.

As before, the worksheet has a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items. As before, each blank is one item, and each cross out is one item. They are all worth 5 points if you attempt them, and an additional 5 points if correct.

We will also give a bonus of 50 points to the three people who get the most items correct. If there is a tie for the top three, we will give the bonus points to all participants who achieve the same number of correct items.

You will have 8 minutes for this round. We do not expect you to finish all the items on the worksheet. We will collect the worksheets at the end of the round to calculate the points. Please turn over the worksheet and begin.

[Visibility treatment]
We are again passing out a worksheet. Please do not look at it yet. In this round you will complete the items on the sheet in front of you, as in the previous round, but the items on this worksheet are different. As before, you will earn 5 points for each item you attempt, and an additional 5 points for each item you complete correctly.

As before, the worksheet has a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items. As before, each blank is one item, and each cross out is one item. They are all worth 5 points if you attempt them, and an additional 5 points if correct.

Recall that your work from this round may be viewed by another participant. Whether or not your work is viewed is going to be determined randomly after you have completed the task for this round. The person who sees your completed work will not be responsible for determining the points you receive, but they will see which items you completed correctly and incorrectly and how many items you did.

Recall that you do not know the chances of your work being reviewed.

You will have 8 minutes for this round. We do not expect you to finish all the items on the worksheet. We will collect the worksheets at the end of the round to calculate the points. Please turn over the worksheet and begin.

[Bonus + visibility treatment]
We are again passing out a worksheet. Please do not look at it yet. In this round you will complete the items on the sheet in front of you, as in the previous round, but the items on this worksheet are different. As before, you will earn 5 points for each item you attempt, and an additional 5 points for each item you complete correctly.

As before, the worksheet has a series of items. The items consist of text extracted from the commercial law for credit, land law and property. You will be asked to fill in the blank for some items and cross out text that is not a part of the law for other items. As before, each blank is one item, and each cross out is one item. They are all worth 5 points if you attempt them, and an additional 5 points if correct.

We will also give a bonus of 50 points to the three people (across both rooms) who get the most items correct. If there is a tie for the top three, we will give the bonus points to all participants who achieve the same number of correct items. We will collect the sheets at the end of the round to calculate the points.

Recall that your work from this round may be viewed by another participant. Whether or not your work is viewed is going to be determined randomly after you have completed the task for this round. The person who sees your completed work will not be responsible for determining the points you receive, but they will see which items you completed correctly and incorrectly and how many items you did. Recall that you do not know the chances of your work being reviewed.

You will have 8 minutes for this round. We do not expect you to finish all the items on the worksheet. We will collect the worksheets at the end of the round to calculate the points. Please turn over the worksheet and begin.

Round 3 *[all groups]*

This is the final round. It consists of an interactive module and a questionnaire. The interactive module in this round is different from what we have done so far. Unlike in other rounds, you are **not** required deal with any commercial law. You are, instead, required to choose between two payment types:

⇒ You can select a payment with certain value.
⇒ Alternatively, you can select a "risky value payment" with an uncertain value.

The value of the risky payment will depend on drawing a ball from a box. The box has some red balls and some black balls, in an unknown proportion. One person from the classroom will draw a ball from the box. If you choose the risky option, your payment will depend on the outcome of the draw. If you choose the certain payment, you will receive that amount regardless of the outcomes of the draw. Relative gains from red or black ball being drawn will change from draw to draw.

Consider the below example, intended to help your understanding. Note this is an example and therefore you will not earn money, although we will ask you to make a choice.

Example:

In this example, the certain value payment is worth 0.50 somani. If you choose this option, let us call it Option (A), then you will get 0.50 somoni for sure for this draw, independent of the color of a ball that is drawn.

The risky option is worth 1 somani if the draw is red and 0.30 somani if the draw is black. If you choose this option, let us call it Option (B), your payment will depend on the color of the ball that is drawn. Note that red balls are always worth more than the certain amount and black balls are always worth less than the certain amount. Note that you need to make your choice before the ball is drawn.

The table below summarizes the options. You need to indicate your choice in the last column.

Option (A)	**Option (B)** - your payment depends on the color of a ball.		**Please indicate in the space below whether you choose option (A) or (B)**
The amount you will get independent of the color of the ball.	if a red is drawn then you will get	if a black is drawn then you will get	
0.50 TJS	1 TJS	0.30 TJS	

We now draw a ball out of the container. If you chose Option (A), your payment would be 0.50 somani, regardless of what color ball was drawn. If you chose option

B, you would have earned [ ] somani because we drew a [ ] ball. This concludes the example.

Round 3a *[all groups]*

For the module, your decisions will determine your actual payment. We will draw a ball 11 times. For each draw, you will decide between Option (A), "Certain value payment", and (B), "Risky value payment", as in the example above. You will make all your choices on the "Decision table" we now pass out. This page contains 11 tables as the above, but with different payment values in Option (B). You will make all your choices before we make any draws. We will collect your "Decision sheets" before making the draws.

After each draw, we will return the ball back to the container. So, across all draws, the chances that a certain color will appear are equal.

Please raise your hand now if you have any questions.

Now, please turn over the sheet in front of you and make your choices.

Round 3b *[all groups]*

Please fill out the questionnaire we now give you.

[Draws from the the box occur here, once all questionnaires are collected.]

End of the Session Discussion *[all groups]*

Our legal expert will now review with you the answers to the items that appeared on the worksheets. You will have the chance to ask any questions about the material throughout this discussion.

[Additional Instructions for Groups 2 and 3 ONLY]

We now randomly select some of your completed worksheets to be available for review. *[Moderator holds up the stack of finished and graded worksheets and pulls randomly from this pile.]* The selected worksheets are available here *[moderator indicates the table where the worksheets are]* for those who would like to see them.

[Time allocated for the discussion. Participants are also given time to review

worksheets and ask questions.]

We will now re-collect the worksheets. It is important that we receive back all worksheets.

Post-discussion *[all groups]*

You will now come up one by one to receive your payment. You will also have the chance to confirm the point totals. You will be asked to sign a receipt of payment, acknowledging that the earnings were from your participation in this interactive session only.

Thank you for your time today.