Social Preferences and Rating Biases in Subjective Performance Evaluations

David Kusterer Dirk Sliwka

June 8, 2022

University of Cologne

Introduction

Subjective Performance Evaluations

- Widely used in firms, educational institutions etc.
- Important when objective criteria to assess performance are unavailable or too costly
- **Observation:** Ratings tend to be **lenient** and **compressed** (e.g. Murphy and Cleveland, 1995, Prendergast, 1999, Moers, 2005)
- Different **purposes** such as allocation of individual bonuses, personnel decisions

This paper

Use standard framework (Prendergast and Topel, 1996) and test implications of

– **performance pay** for worker
– **accuracy incentives** for supervisor
– **signal precision**
– supervisor social preferences

on rating leniency, compression, and rating errors in an incentivized experiment

Subjective Performance Evaluations

- Widely used in firms, educational institutions etc.
- Important when objective criteria to assess performance are unavailable or too costly
- **Observation:** Ratings tend to be **lenient** and **compressed** (e.g. Murphy and Cleveland, 1995, Prendergast, 1999, Moers, 2005)
- Different **purposes** such as allocation of individual bonuses, personnel decisions

This paper

Use standard framework (Prendergast and Topel, 1996) and test implications of

	Affected measure	Prediction	Data
– **performance pay** for worker	avg. rating	↑	↑
– **accuracy incentives** for supervisor	avg. rating	↓	↓
– **signal precision**	compression	↓	↓
– supervisor social preferences	avg. rating	↑	depends

on rating leniency, compression, and rating errors in an incentivized experiment

Literature

- **Effect of ratings on behavior**
  Imposing a forced distribution raises worker performance when these workers work separately (Berger et al., 2013). Workers negatively reciprocate low ratings, these reactions depend on the worker's over- and underconfidence (Sebald and Walzl, 2014, Bellemare and Sebald, 2019).
- **Uncertainty in performance signal**
  Less precise signals cause higher rating compression with ambiguous effects on rating leniency (see e.g. Bol and Smith, 2011 and Bol et al., 2016). In feedback provision on online markets, uncertainty about the cause of quality deficiencies increases rating leniency and compression (Rice, 2012, Bolton et al., 2019).
- **Social preferences**
  Breuer et al. (2013) find evidence that supervisors tend to assign better ratings at the same level of objective performance to workers with whom they have worked for a longer time before.
- **Calibration committees (provision of accuracy incentives)**
  Calibration committees reduce leniency (Demeré et al., 2019) by disciplining supervisors (Grabner et al., 2020). Performance evaluations by multiple raters provide more accurate ratings (Ockenfels et al., 2020).
- **Preference for leniency**
  In public goods games, monitoring with severity errors decreases contributions more than one with leniency errors (Dickson et al., 2009), subjects have a larger willingness to pay to play in an environment that makes leniency errors compared to severity errors (Markussen et al., 2016). In a principal-agent environment, a monitoring technology that creates leniency errors decreases effort by an agent less than one with severity errors (Marchegiani et al., 2016).

A Simple Model

Model Setup (based on Prendergast and Topel, 1996)

- A **supervisor** evaluates the performance of a **worker**
- Supervisor observes vector s of $i = 1, \dots, n$ **performance signals** $s_i = a + \epsilon_i$, where $a \sim N(m, \sigma_a^2)$ is the agent's true performance and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ are noise terms
- Supervisors determines a **performance rating** r
- **Worker's utility** is

$$\alpha + (\beta + b)r,$$

where α is a fixed payment, β is the worker's monetary and b the worker's psychological utility from a rating

- **Supervisor's utility** is

$$\eta \underbrace{[\alpha + (\beta + b)r]}_{\text{Worker's utility}} - \frac{\gamma + \lambda}{2} \underbrace{E\left[(r - a)^2 | s\right]}_{\text{Rating error}^2},$$

where η measure social preferences, γ is intrinsic preference for accuracy and λ is material incentive to rate accurately

- Supervisor chooses r to maximize her utility
- Signal average $\bar{s} = \frac{1}{n}\sum_{i=1}^{n} s_i$ is sufficient statistic for estimating a

Proposition 1

After having observed performance signal $\bar{s}$ the supervisor reports

$$r(\bar{s}) = \underbrace{\frac{\eta(\beta+b)}{\gamma+\lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2+\sigma_\varepsilon^2}\cdot m}_{\textit{Intercept}} + \underbrace{\frac{n\sigma_a^2}{n\sigma_a^2+\sigma_\varepsilon^2}}_{\textit{Slope}}\cdot\bar{s}.$$

Hypotheses: Ratings

$$r(\bar{s}) = \underbrace{\frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m}_{\text{Intercept}} + \underbrace{\frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2}}_{\text{Slope}} \cdot \bar{s}$$

Hypothesis 1: Performance pay (β)

Bonus payments lead to higher rating leniency.

Hypothesis 2: Accuracy pay (λ)

A reward for accuracy reduces rating leniency. This reduction in leniency will be larger when the agent receives performance pay.

Hypothesis 3: More signals (n)

If the supervisor observes more signals, the slope of the rating function increases, i.e. $\frac{\partial r}{\partial \bar{s} \partial n} > 0$ *and its intercept decreases* $\frac{\partial r}{\partial n}\Big|_{\bar{s}=0} < 0$.

Hypothesis 4: Social preferences (η)

Rating leniency is higher when supervisors have stronger social preferences. This effect is stronger when ratings determine bonus payments.

Experimental Design

Structure of the Experiment

Part 1 (Workers)

- **Main task:** Enter 10 captchas on 10 screens with varying time limits (“Entry Task”)

Part 2 (Supervisors)

- **Main task:** Submit a rating for one worker (“Rating Task”)
- Practice the Entry Task
- Learn about performance of all workers (mean, sd, histogram)
- Signal: sample of the 10 screens from one worker
- Social Value Orientation Task (SVO) towards a random worker (not the matched one) as measure of social preferences

Part 3 (Workers)

- Learn their performance and their rating
- Learn bonus and treatment
- **Main task:** SVO towards supervisor who rated them

Treatments

Treatment variables

- **Performance pay** for workers β
- **Incentives for accuracy** for supervisor λ
- **Number of signals** supervisor receives n

- Supervisor payoff: $\$4 - \lambda(r-a)^2 + 0.01a$
- Worker payoff: $\$1 + \beta r/100$

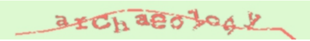
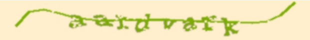
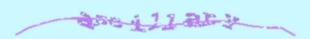
Six treatments

Name	β	λ	n
NP-NA-S1	\$0	0	1
NP-A-S1	\$0	0.004	1
P-NA-S1	\$2	0	1
P-A-S1	\$2	0.004	1
P-NA-S4	\$2	0	4
P-A-S4	\$2	0.004	4

Entry Task Page 1

Time left to complete this page: **0:17**

Please enter the text shown in the images. All images contain real words.

Image	Entry
	
	
	
	
	
	
	
	
	
	

Next

Rating Task

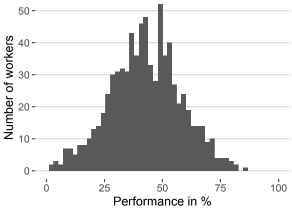
Show Instructions

Distribution of performance

The **average performance** of 780 workers who completed the Entry Task is **42.5%**. This means that on average, they entered the text correctly on 42.5 of the 100 images. The **standard deviation** (a measure of how far the performance is spread out) of these workers is **15.5**.

The graph below shows the distribution of performance of 780 workers who completed the Entry Task.

You can read it in the following way: For each level of performance (on the axis at the bottom), it shows the number of workers with that performance (on the axis at the left).



Your worker's performance

The worker matched to you had the following performance on 1 out of 10 pages that was randomly selected. Note that their performance on the other 9 pages will not be revealed to you.

Remember that the 10 pages had different time limits such that the revealed performance can be from a page with any of the time limits mentioned in the Instructions (17, 19, 21, 23, or 25 seconds).

Page	Number of correct entries
1	
2	
3	**5**
4	
5	
6	
7	
8	
9	
10	

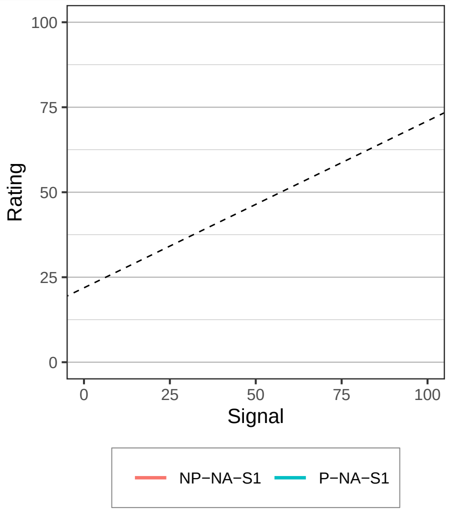
Results

Descriptives

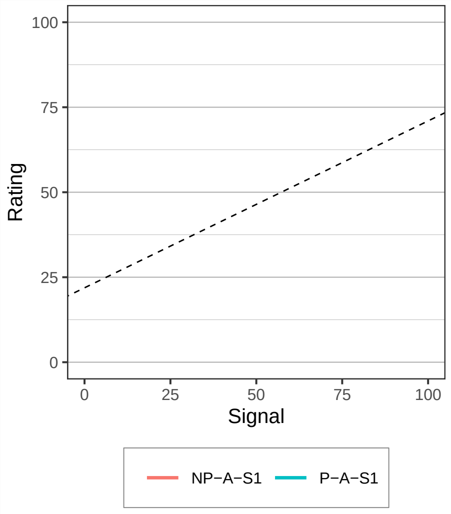
$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s}$$

$$\frac{\partial r}{\partial \beta} > 0 \qquad \frac{\partial^2 r}{\partial s \partial \beta} = 0 \qquad \frac{\partial r}{\partial \lambda} < 0 \qquad \frac{\partial^2 r}{\partial \lambda \partial \beta} < 0$$

No accuracy pay Regressions



Accuracy pay Regressions

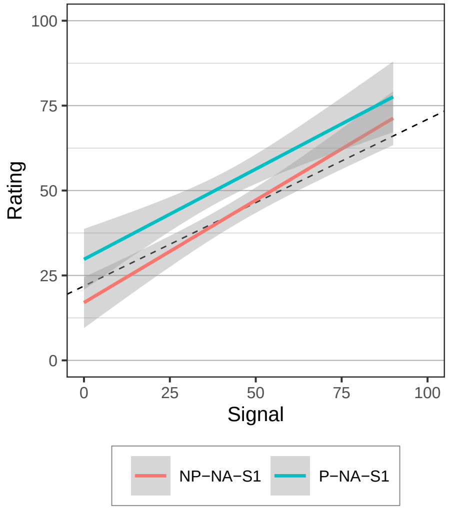


$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s}$$

$$\frac{\partial r}{\partial \beta} > 0 \qquad \frac{\partial^2 r}{\partial s \partial \beta} = 0 \qquad \frac{\partial r}{\partial \lambda} < 0 \qquad \frac{\partial^2 r}{\partial \lambda \partial \beta} < 0$$

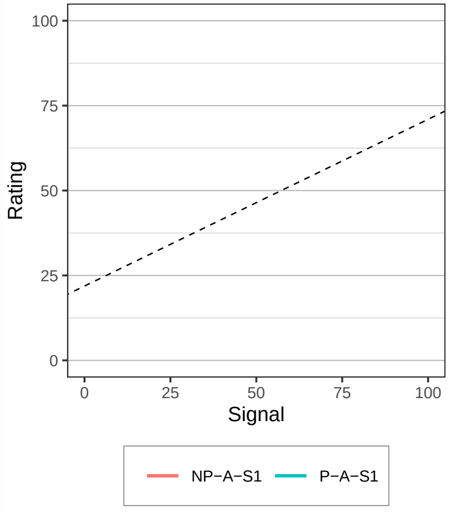
No accuracy pay

Regressions



Accuracy pay

Regressions



Performance Pay and Incentives for Accuracy

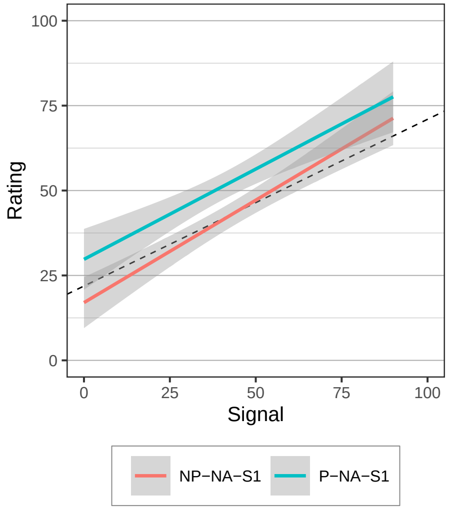
Descriptives

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s}$$

$$\frac{\partial r}{\partial \beta} > 0 \qquad \frac{\partial^2 r}{\partial s \partial \beta} = 0 \qquad \frac{\partial r}{\partial \lambda} < 0 \qquad \frac{\partial^2 r}{\partial \lambda \partial \beta} < 0$$

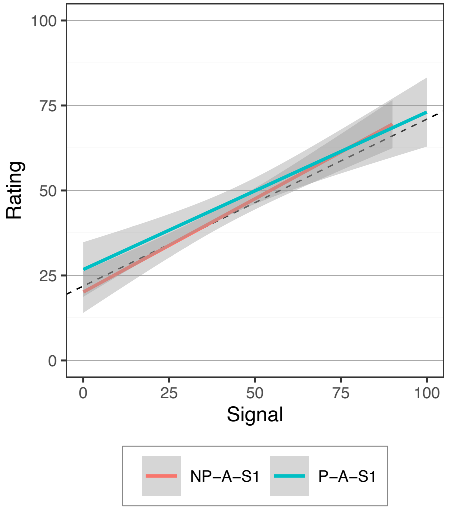
No accuracy pay

Regressions



Accuracy pay

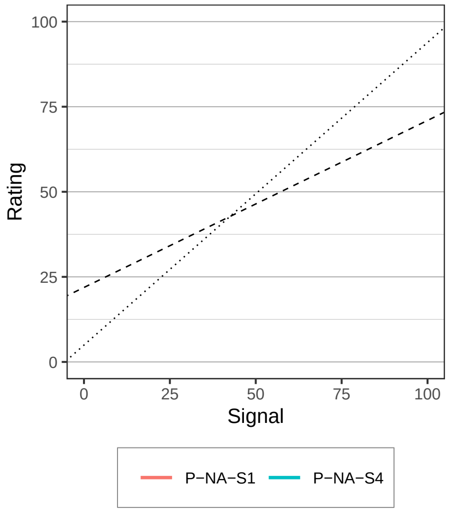
Regressions



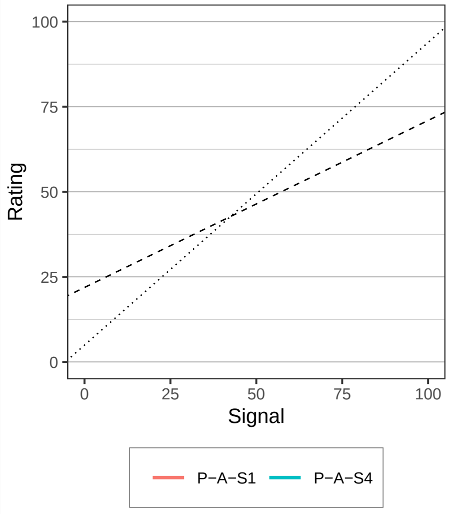
Signal Precision and Compression

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial^2 r}{\partial s \partial n} > 0 \qquad \left.\frac{\partial r}{\partial n}\right|_{s=0} < 0$$

No accuracy pay



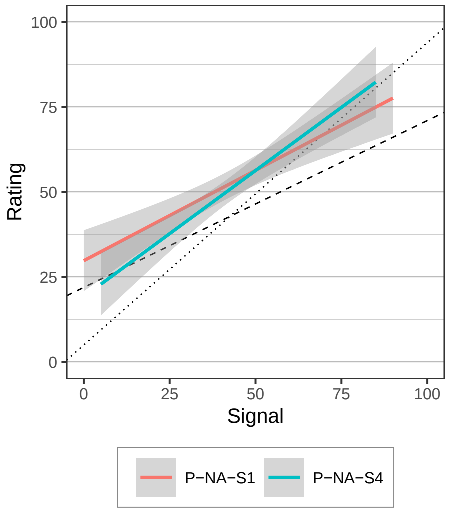
Accuracy pay



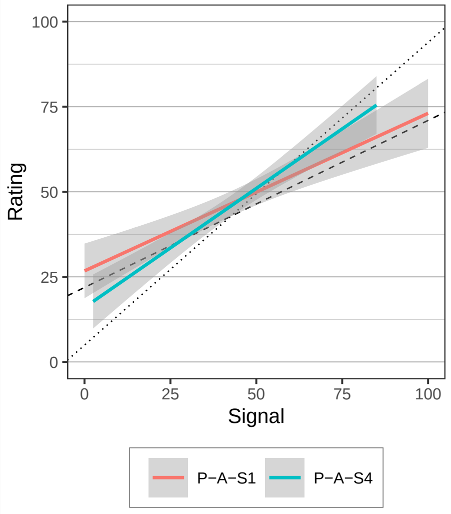
Signal Precision and Compression

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial^2 r}{\partial s \partial n} > 0 \qquad \left.\frac{\partial r}{\partial n}\right|_{s=0} < 0$$

No accuracy pay



Accuracy pay



Recap of main hypotheses

Hypothesis 1: Performance pay (β)

Bonus payments lead to higher rating leniency.

Result: Introducing performance pay leads to more lenient ratings.

Hypothesis 2: Accuracy pay (λ)

A reward for accuracy reduces rating leniency. This reduction in leniency will be larger when the agent receives performance pay.

Result: Accuracy pay only reduces leniency when there is performance pay.

Hypothesis 3: More signals (n)

If the supervisor observes more signals, the slope of the rating function increases, i.e. $\frac{\partial r}{\partial \bar{s} \partial n} > 0$ and its intercept decreases $\frac{\partial r}{\partial n}\Big|_{\bar{s}=0} < 0$.

Result: More signals lead to a lower intercept and a larger slope such that ratings vary more with the signal.

Supervisors' Social Preferences

Figure | Distribution of SVO angle

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial r}{\partial \eta} > 0 \qquad \frac{\partial^2 r}{\partial \eta \partial \beta} > 0 \qquad \frac{\partial^2 r}{\partial \eta \partial \lambda} < 0$$

	Rating (No acc. inc.)	Rating (Acc. inc.)	Rating (pooled)	Sq. rating error (pooled)
Prosocial	8.737**	0.334	-6.876	-147.8*
	(3.633)	(2.936)	(4.191)	(82.74)
Performance pay	12.71***	5.947	6.180***	147.6*
	(4.294)	(3.959)	(1.789)	(78.88)
Prosocial × Performance pay	-5.956	-4.867		
	(5.457)	(4.929)		
Signal average	0.569***	0.505***	0.413***	
	(0.0636)	(0.0564)	(0.0744)	
Prosocial × Signal average			0.212**	
			(0.0888)	
Accuracy pay			-2.569	-139.2*
			(1.799)	(79.24)
Constant	13.93***	21.73***	25.60***	644.7***
	(3.312)	(3.022)	(3.307)	(69.80)
Observations	260	260	520	520
R^2	0.290	0.270	0.277	0.020

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The Variance of Ratings and the Notion of Rating Compression

Rating variance is

$$V[r] = \left(\frac{\beta + b}{\gamma + \lambda}\right)^2 \sigma_\eta^2 + \frac{\sigma_a^4}{\sigma_a^2 + \frac{\sigma_\varepsilon^2}{n}} + \sigma_\zeta^2, \qquad \zeta \sim N\left(0, \sigma_\zeta^2\right)$$

Observations

- Variance often used in field studies to assess compression
- When n is larger, rating variance is larger ($\frac{\partial V[r]}{\partial n} > 0$)

$\Rightarrow$ More precise signal goes along with less “rating compression” or “centrality bias”

The Variance of Ratings and the Notion of Rating Compression

Rating variance is

$$V[r] = \left(\frac{\beta + b}{\gamma + \lambda}\right)^2 \sigma_\eta^2 + \frac{\sigma_a^4}{\sigma_a^2 + \frac{\sigma_\varepsilon^2}{n}} + \sigma_\zeta^2, \qquad \zeta \sim N\left(0, \sigma_\zeta^2\right)$$

Observations

- Variance often used in field studies to assess compression
- When n is larger, rating variance is larger ($\frac{\partial V[r]}{\partial n} > 0$)

⇒ More precise signal goes along with less “rating compression” or “centrality bias”

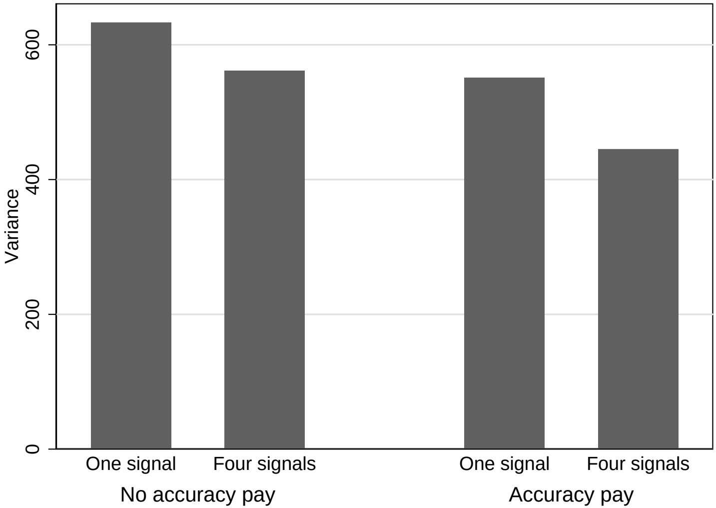
- We find the opposite: Variance decreases with more signals (although compression decreases)

⇒ Caution when using variance as proxy for rating compression

- Speculation: Supervisors economize on cognitive costs of information processing?

The Variance of Ratings and the Notion of Rating Compression

Regression



Conclusion

Conclusion

Canonical model organizes data well

- **Performance pay** for worker (ratings ↑)
- **Accuracy incentives** for supervisor (ratings ↓)
- **Signal precision** (compression ↓)

Implications

- **Social preferences** have complex relation with ratings: prosocial supervisors are sometimes more lenient, but also tend to give more accurate ratings
- Caution when using **rating variance** as proxy for rating compression
- **Purpose of rating:** Providing accuracy incentives more important when bonus is tied to the ratings

Thank you for your attention!

Additional results

Workers' reactions to ratings

	(1) Perf. inc.	(2) No perf. inc.	(3) Pooled	(4) Pooled
Rating deviation	0.164***	0.131***	0.125***	0.201***
	(0.0315)	(0.0432)	(0.0423)	(0.0532)
Actual performance	-0.0130	0.0292	0.000823	0.00244
	(0.0464)	(0.0586)	(0.0366)	(0.0366)
Rating dev. × Performance pay			0.0422	
			(0.0513)	
Performance pay			1.128	1.169
			(1.158)	(1.146)
$\max\{\text{Rating deviation}, 0\}$				-0.0772
				(0.0757)
Constant	19.49***	16.55***	17.75***	18.34***
	(2.159)	(2.589)	(1.777)	(1.864)
Observations	510	254	764	764
R^2	0.057	0.036	0.054	0.055

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Additional slides

Sketch of Proof of Proposition 1

Back

Supervisor maximizes

$$\max_r \eta \cdot (\alpha + (\beta + b) \cdot r) - \frac{\gamma + \lambda}{2} E\left[(r-a)^2 \middle| \bar{s}\right].$$

As $E\left[(r-a)^2 \middle| \bar{s}\right] = V[r-a|\bar{s}] + (E[r-a|\bar{s}])^2$ and $V[r-a|\bar{s}] = V[a|\bar{s}] = V[a] - \frac{(Cov[a,\bar{s}])^2}{V[\bar{s}]}$, we have

$$E\left[(r-a)^2 \middle| \bar{s}\right] = \frac{\sigma_a^2 \sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} + \left(r - m - \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2}(\bar{s} - m)\right)^2.$$

FOC of supervisor's problem is

$$\eta(\beta + b) - (\gamma + \lambda)\left(r - m - \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2}(\bar{s} - m)\right) = 0$$

from which we obtain

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s}.$$

Hypotheses: Rating errors

Proposition 2

The expected squared rating error is

$$E\left[(r-a)^2\right] = \frac{\sigma_a^2\sigma_\varepsilon^2}{n\sigma_a^2+\sigma_\varepsilon^2} + \frac{(\beta+b)^2}{(\gamma+\lambda)^2}\left(\sigma_\eta^2+m_\eta^2\right).$$

Hypothesis 5: Rating errors

Rating errors are larger when agents receive bonus payments and smaller when supervisors are rewarded for accuracy and when they observe more performance signals. A reward for accuracy reduces rating errors to a larger extent when ratings determine agents' bonus payments.

Sketch of Proof of Proposition 2

Back

Substitute

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s}.$$

into

$$E\left[(r-a)^2 \middle| \bar{s}\right] = \frac{\sigma_a^2 \sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} + \left(r - m - \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2}(\bar{s} - m)\right)^2.$$

and using $E\left[\eta^2\right] = V[\eta] + E[\eta]^2$ we obtain

$$\frac{\sigma_a^2 \sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} + \left(\frac{\eta(\beta + b)}{\gamma + \lambda}\right)^2$$

from which follows that

$$E\left[(r-a)^2\right] = \frac{\sigma_a^2 \sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} + \frac{(\beta + b)^2}{(\gamma + \lambda)^2}\left(\sigma_\eta^2 + m_\eta^2\right)$$

Procedures

- Experiment conducted on Amazon **MTurk** with subjects from the US
- Implemented in **oTree** (Chen et al., 2016)
- **Treatment assignment** after gathering data in Part 1 to ensure similar performance distribution across treatments
- **Comprehension questions:** 3 attempts
- **MTurk qualifications:** 1000 HITs, 98% approval rate
- 780 workers, 780 supervisors, 130 of each type per treatment
- **Duration:**
  - Part 1: 11 minutes
  - Part 2: 16 minutes
  - Part 3: 4 minutes
- 50.5% female, on average 37.9 years

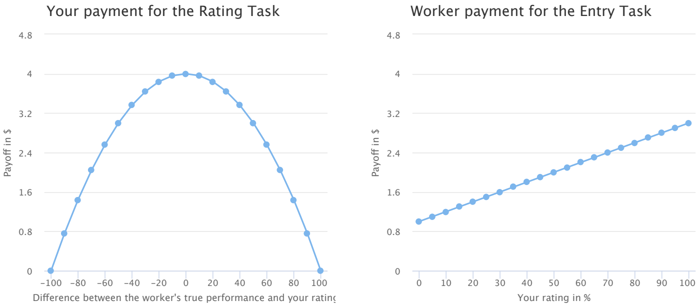
The payoffs

Your payment:

- For the Rating Task you receive \$4.00 – 0.9 × (true performance - rating)^2/2250, but not less than \$0.00.
  The payment will be the higher, the closer your rating is to the true performance (see figure below).
- You will also receive \$0.01 for every image the worker matched to you entered correctly over all 10 pages.
  For example, if they entered 0 images correctly, you receive \$0.00, if they entered 50 images correctly, you receive \$0.50, and if they entered 100 images correctly, you receive \$1.00. (This payment does not depend on your rating, only on the worker's actual performance.)

The worker receives a payment of \$1.00 + \$2.00 × (**your rating**)/100.
The worker's payment increases in your rating (see figure below). The higher the rating you give, the higher the worker's payment will be. (The worker's payment is paid by us and not deducted from your earnings.)

These graphs illustrate your payment for the Rating Task and the payment of the worker matched to you based on the rating you give and their true performance:



Treatments

Your rating

How would you rate the worker?

As a guidance, ratings should reflect the percentage of correctly entered images by the worker.

%

Given your currently entered rating, the worker would receive a bonus of $--.

Descriptives

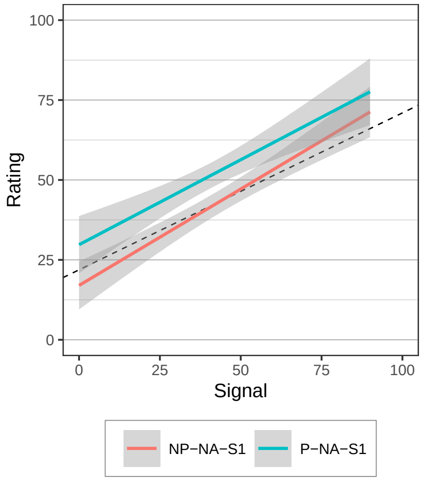
Back

	Performance		Rating	
Treatment	Mean	Variance	Mean	Variance
NP-NA-S1	42.55	242.56	43.22	638.29
P-NA-S1	42.79	242.69	51.49	633.26
P-A-S1	42.44	242.51	46.66	551.06
NP-A-S1	42.37	241.27	42.58	426.60
P-NA-S4	42.66	240.61	50.42	561.55
P-A-S4	42.25	242.05	45.55	445.34

Performance Pay and Rating Leniency

Back

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial r}{\partial \beta} > 0 \qquad \frac{\partial^2 r}{\partial s \partial \beta} = 0$$



	(1) Rating	(2) Rating	(3) Rating
Performance pay	8.277***	9.732***	12.74**
	(3.127)	(2.672)	(6.268)
Signal		0.573***	0.603***
		(0.0653)	(0.0863)
Signal × Performance pay			-0.0717
			(0.132)
Constant	43.22***	18.30***	17.00***
	(2.216)	(3.096)	(3.778)
Observations	260	260	260
R^2	0.026	0.274	0.275

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Signal precision and Compression

Back

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial^2 r}{\partial s \partial n} > 0 \qquad \left.\frac{\partial r}{\partial n}\right|_{s=0} < 0$$

	(1)	(2)	(3)	(4)
Signal average	0.577***	0.494***	0.494***	0.532***
	(0.0510)	(0.0688)	(0.0689)	(0.100)
Four signals	-1.214	-10.76**	-11.19**	-10.62
	(1.814)	(4.693)	(5.106)	(7.110)
Accuracy pay	-5.475***	-5.394***	-5.857**	-2.992
	(1.814)	(1.806)	(2.723)	(6.843)
Four signals × Signal average		0.227**	0.226**	0.211
		(0.0969)	(0.0970)	(0.148)
Accuracy pay × Four signals			0.925	-0.118
			(3.615)	(9.429)
Accuracy pay × Signal average				-0.0684
				(0.138)
Accuracy pay × Four signals × Signal average				0.0252
				(0.195)
Constant	27.60***	31.06***	31.27***	29.74***
	(2.884)	(3.560)	(3.722)	(5.002)
Observations	520	520	520	520
R^2	0.227	0.234	0.235	0.235

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

More signals: Effect on Average Rating

Back

	(1) No acc. inc.	(2) Acc. inc.	(3) Pooled
Signal average	0.612*** (0.0755)	0.548*** (0.0688)	0.577*** (0.0510)
Four signals	-1.810 (2.696)	-0.683 (2.433)	-1.214 (1.814)
Accuracy pay			-5.475*** (1.814)
Constant	26.46*** (4.005)	23.09*** (3.655)	27.60*** (2.884)
Observations	260	260	520
R^2	0.207	0.233	0.227

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Incentives for Accuracy and Agents' Performance Pay

Back

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial r}{\partial \lambda} < 0 \qquad \frac{\partial^2 r}{\partial \lambda \partial \beta} < 0$$

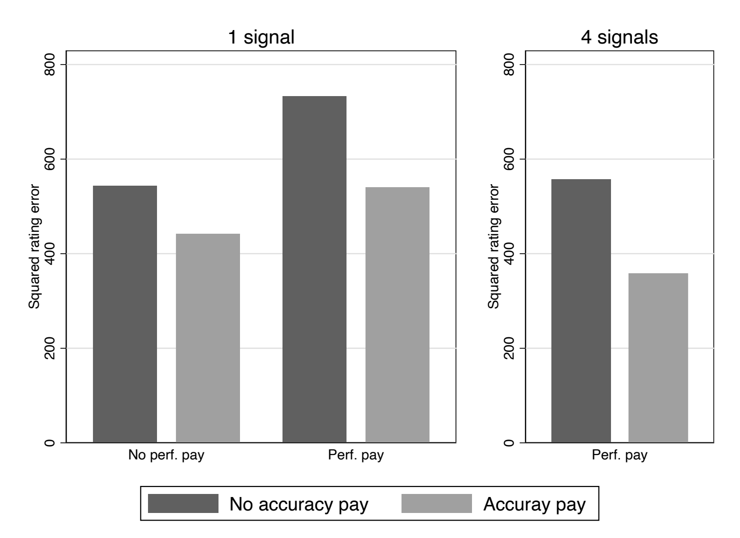
	(1) No perf. inc.	(2) Perf. inc.	(3) Pooled	(4) Pooled
Signal	0.578***	0.494***	0.540***	0.606***
	(0.0541)	(0.0689)	(0.0431)	(0.0750)
Accuracy pay	0.837	-5.857**	0.739	3.395
	(2.325)	(2.723)	(2.332)	(4.084)
Performance pay			9.647***	13.04***
			(2.688)	(4.576)
Performance pay × Accuracy pay			-6.690*	-6.708*
			(3.581)	(3.578)
Signal × Accuracy pay				-0.0608
				(0.0850)
Signal × Performance pay				-0.0789
				(0.0864)
Constant	18.09***	31.27***	19.77***	16.87***
	(2.681)	(3.722)	(2.441)	(3.392)
Observations	260	260	520	520
R^2	0.335	0.194	0.271	0.274

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Rating errors

Main results

$$E\left[(r-a)^2\right] = \frac{\sigma_a^2\sigma_\varepsilon^2}{n\sigma_a^2+\sigma_\varepsilon^2} + \frac{(\beta+b)^2}{(\gamma+\lambda)^2}\left(\sigma_\eta^2+m_\eta^2\right) \qquad \frac{\partial E[(r-a)^2]}{\partial\lambda}<0 \qquad \frac{\partial E[(r-a)^2]}{\partial n}<0 \qquad \frac{\partial E[(r-a)^2]}{\partial\beta}>0$$



Rating errors

$$E\left[(r-a)^2\right] = \frac{\sigma_a^2\sigma_\varepsilon^2}{n\sigma_a^2+\sigma_\varepsilon^2} + \frac{(\beta+b)^2}{(\gamma+\lambda)^2}\left(\sigma_\eta^2+m_\eta^2\right) \qquad \frac{\partial E[(r-a)^2]}{\partial \lambda} < 0 \qquad \frac{\partial E[(r-a)^2]}{\partial n} < 0 \qquad \frac{\partial E[(r-a)^2]}{\partial \beta} > 0$$

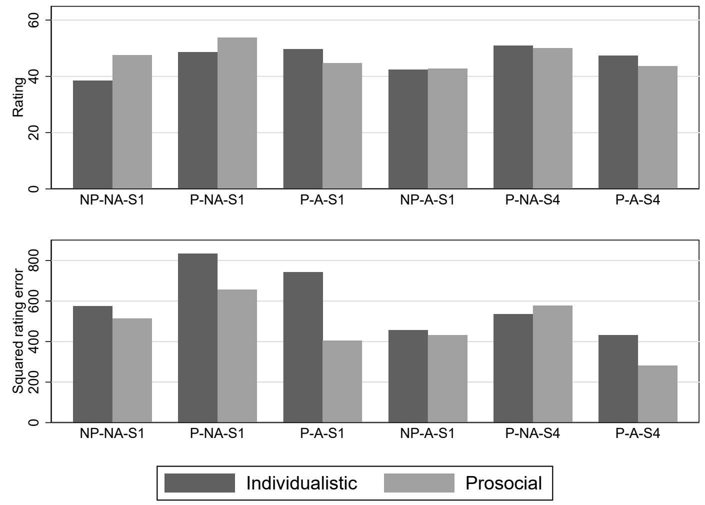
	(1)	(2)
Performance pay	143.6*	190.9*
	(78.43)	(103.7)
Accuracy pay	-164.1**	-101.0
	(64.07)	(92.03)
Four signals	-178.4**	-178.4**
	(84.36)	(84.40)
Performance pay × Accuracy pay		-94.59
		(124.9)
Constant	574.7***	543.2***
	(57.71)	(69.24)
Observations	780	780
R^2	0.016	0.016

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Supervisors' Social Preferences

Regression

$$r(\bar{s}) = \frac{\eta(\beta + b)}{\gamma + \lambda} + \frac{\sigma_\varepsilon^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot m + \frac{n\sigma_a^2}{n\sigma_a^2 + \sigma_\varepsilon^2} \cdot \bar{s} \qquad \frac{\partial r}{\partial \eta} > 0$$

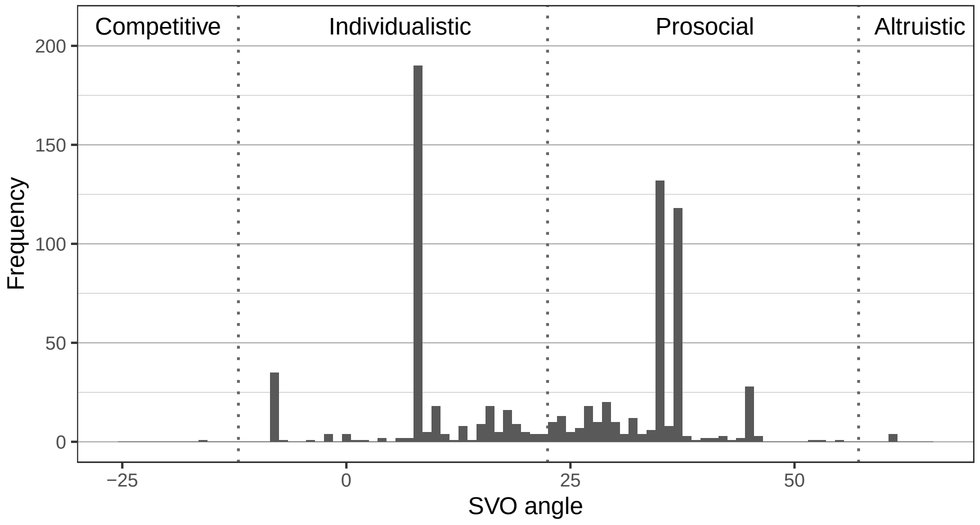


Social value orientation

- Measures subjects' preferences over a set of allocations of money between themselves and another subject.
- **Slider measure (Murphy et al., 2011)**, 6 dictator games.
- Measure: Angle between own payoffs and other's payoffs.
- Run after main tasks.
- **Part 2:** Randomly matched with another worker, but not the rated one.
- **Part 3:** Matched with supervisor who gave rating.
- Four types:
  1. **Individualistic**: Maximize own payoff
  2. **Prosocial**: Maximize joint surplus and/or reduce payoff inequality
  3. Competitive: Maximize payoff difference, with themselves ahead
  4. Altruistic: Maximize the other subject's payoff
- Types 3 and 4 empirically less relevant, merged into 1 and 2.

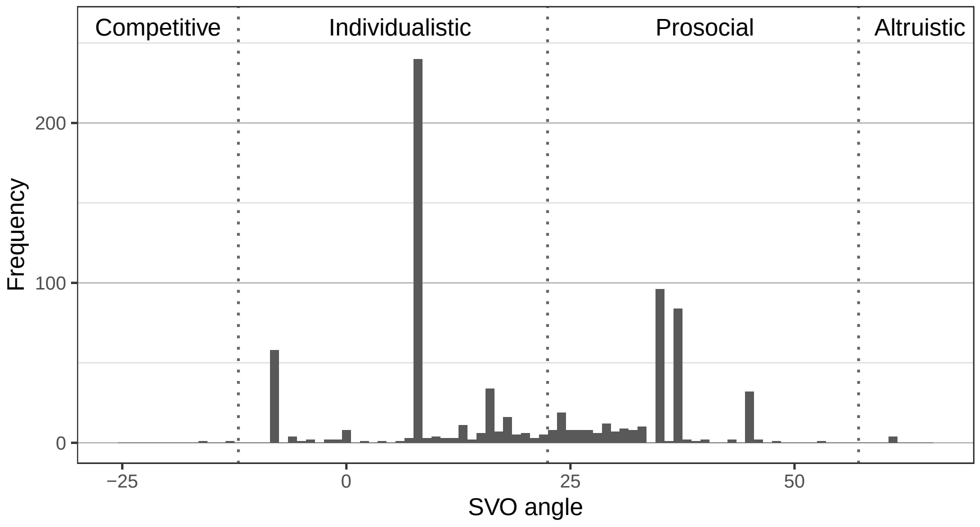
Supervisor SVO angle distribution

Back



Worker SVO angle distribution

Back



The Variance of Ratings and the Notion of Rating Compression

Figure

	(1) P-NA-S1	(2) P-NA-S4	(3) P-A-S1	(4) P-A-S4
Signal average	5.315***	7.429***	4.631***	6.997***
	(0.993)	(1.148)	(0.839)	(0.932)
Constant	29.74***	19.12***	26.75***	16.01***
	(4.530)	(5.167)	(4.058)	(4.229)
Observations	130	130	130	130
Variance of dependent variable	633.3	561.5	551.1	445.3
Variance of predicted values	115.8	138.4	105.9	136.1
Variance of residuals	517.4	423.2	445.1	309.2
R^2	0.183	0.246	0.192	0.306

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$