Fairness Preferences in the face of Limited Information. Pre-analysis plan for new spectator sessions September 2017.

By Caroline Bonn, Alexander Cappelen, Bertil Tungodden and Thomas de Haan.

September 20, 2017

Introduction

In this project we investigate, theoretically and experimentally, the role of limited information about performance in a collective production task on redistributional preferences.

In our study we focus both on distribution decisions by one of two matched workers, as well as distribution decisions made by a monetarily unaffected spectator. For both the setting with a stakeholder and a spectator redistribution decision, we will compare two treatments. One, where decision makers have full information regarding what influences worker performance and what influence 'random factor' luck had in determining the workers' contribution to the total income. Two, where decision makers have limited information regarding the worker's actual performance.

Building on the fairness preferences model from Cappelen et al. (2007, AER), we predict that stakeholders who are ex-ante pessimistic about other workers' performance, should become less prone to redistribute in the face of limited information if they hold a meritocratic fairness ideal. In contrast meritocratic spectators are predicted to become more egalitarian in their redistribution choices when facing limited information.

To test our hypotheses regarding distributional preferences in the face of limited information, both in a stakeholder and spectator setting, we ran experimental sessions at the University of Bergen and and the Norwegian School of Economics in Bergen in April and May 2016. [1] Our results for the stakeholder setting can be summarized by stakeholders sharing on average equal amounts in the Full Information and Limited Information treatments. We do see, both in the sessions from 2016 and May 2017, that there is a substantial amount of pluriformity in the amounts that spectators share and apparent decision rules that spectators

[1] Our pre-analysis plan was uploaded to the AEA rct registry before conducting the experiment in 2016 and can be found here: https://www.socialscienceregistry.org/trials/1151 (for the second experiment in 2017, which only featured a minor departure from original design in terms of the random factor, no extra pre-analysis plan was uploaded).

take. This much in line with earlier findings and theoretical predictions from Cappelen et al (2007,2010) and Almås et al (2016).

Results for the spectator treatments for the sessions conducted in 2016 (the sessions in May 2017 only consisted of stakeholder treatments) revealed no significant difference in the levels of implemented equality by the spectators. For these sessions however we used a relatively low level of noise compared to the variance in worker performance and with 35 spectator decisions per treatment, our power in this first session for the specators was relatively low.

We did find some evidence, via the elicitation of beliefs, that many specators in the 2016 sessions only partially updated given a noisy production signal, in line with our theory. This has led us to decide to run a new round of sessions late this September and give the proposed effect for the spectators a chance with a sufficiently powered experiment.

Experimental design

The experiment will consist of two parts. In the first part workers recruited via Amazon Mturk will work on a real effort task. In the second part, spectators recruited from the first years student population at the NHH Norwegian School of Economics will decide on how certain amounts of bonus payments will be distributed over two randomly paired Mturk workers from the first part of the experiment. The first part: Real effort task conducted by Mturk workers. Workers will be recruited via the Amazon Mturk account of the Choice Lab research group. The task the workers will perform is a qualtrix implementation of the 'encription task' introduced for experiments by Benndorf, V., Rau, H.A., and Sölch, C. (2014). In this task workers are shown three letters and at the bottom of the screen all letters of the alphabet, in random order, with below each letter a random number between 100 and 1000. For the three letters shown, the workers must fill out the corresponding number. After this, they can continue to the next screen where they will receive a new three letters. Workers will have exactly 15 minutes to work on the task. Workers will earn points depending on how many times they correctly filled out the three codes on a screen. Next to this, for each worker individually, an integer 'random factor' will be drawn from a normal distribution with mean 0 and standard deviation of 20. The sum of the worker's task performance and the random factor will add to be the 'production' that each Mturk worker will bring to the second part of the experiment. Each Mturker who finishes at least 20 will earn 2 dollar as a fixed fee and workers earn a bonus depending on a number of factors being; their performance and random factor, a random other worker's performance and random factor and the distribution decision made by a spectator. The Mturk workers are informed about this procedure.

The second part: Distributional choices by spectators. The second part of the design will be conducted at the Norwegian School of economics on the 27th 28th and 29th of September.

Students from the first year course "Business Ethics" will be recruited via the course.[2] They will be instructed that the experiment results will be used for actual research purposes, that they will receive a payment for participating in the experiment, which will depend in part on their decisions, and the experiment will take place in a separate time and location compared to the classes for the course. The participants will be asked to perform a number of tasks.

Experiment sequence

- Participants will read instructions and answer a comprehension questionnaire
- Participants will answer incentivized questions regarding two Bayesian updating scenario's, a 'disease test' scenario and the 'taxi cab color' scenario. We expect participants updating with a high degree of 'base-rate neglect' to show a quantitatively (and possibly) qualitatively different treatment effect. This task will pre-classify participants along the base-rate weighting dimension.
- Participants are then informed about the encryption task Mturk workers performed in the week before the experiment. The participants will be shown a histogram distribution of the amount of correct encriptions the different workers managed to complete in the given time.
- Participants will then receive information on the productions (in the each treatment) and performances and random factors (in the Full Information treatment) regarding a set of 10 randomly chosen pair of Mturk workers from the first part of the experiment. For each pair, the participants will be asked to make a distributional decision, splitting the total production points over the two Mturk workers.
- Participants in the Limited Information treatment are then elicited their beliefs regarding the individual performances of the 10 pairs they made a decision for.
- After this participants will make a number of further, non-incentivized choices that will be used for purposes of the course.
- The Participants will fill out a post-experimental questionnaire which will include questions on background variables (age, gender, political preferences). After this participants will be payed out and leave the experiment.

We plan to recruit around 800 workers on Mturk to work on the task to match with the approximately 450 expected participants for the spectator decisions.

[2] All first year students at the NHH are expected to follow the Business Ethics course and participation in the experiment is made a compulsory part of the course.

Hypotheses and analysis strategy

The planned experiment will consist of 2 treatments, the Full Information and Limited Information treatment. Our main hypothesis for the experiment is that we expect the spectators to implement less unequal splits in the Limited Information treatment compared to the Full Information treatment. The main way we plan to test this hypothesis is to take both the average absolute and average relative inequality each spectator has implemented over the 10 distribution choices and compare treatments, both using a Mann-Whitney test and linear regressions using background controls, to test for an effect.

Control variables: Age, gender, political preference, and the results from the two Bayesian updating task we elicit before the distribution decision.

Heterogeneity: As our hypothesized treatment effect is dependent on the way the spectators update we expect to only observe a different treatment effect for those participants that do not reveal themselves as a 'base rate neglect' type in the Bayesian task. We will test this by interacting the treatment effect with the type classification from the Bayesian tasks.

Theoretical prediction

Our main reason to hypothesize a reduced level of implemented inequality stems from the observation from previous fairness distribution studies that a substantial part of decision-makers tend to follow a meritocratic principle when it comes to inequality (Cappelen et al, 2010, Almås et al, 2016). Meritocrats implement a level of inequality that is proportional to the difference in performance in case of Full Information, and proportional to the difference in expected performance in the Limited Information setting. So we need to compare the average difference in productivity $E\left(|p_1 - p_2|\right)$ for the Full Information case, with the average difference in expected productivity $E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right)$ for the Limited Information case. We assume that each of the two worker's performances is drawn from a normal distribution with mean μ_p and standard deviation σ_p. Next to this, for each worker a random factor ε is drawn, individually, from a normal distribution with mean 0 and standard devation σ_ε. We can then derive that $E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right) < E\left(|p_1 - p_2|\right)$ if $\sigma_\varepsilon > 0$. The derivation is added in the appendix below.

Power analysis

To approximate the probability that we will find an effect, we will use choice simulations using a calibrated version of the model used by Cappelen et al. (2007 & 2010). For the worker performance, we draw each simulated worker's performance from a normal distribution with mean 60 and standard deviation 20.[3] The random factor is also normally distributed with

[3] This based on an early tryout of the real effort task on Mturk

a mean of 0 and standard deviation of (20. For the spectators, we randomly draw a norm type. With 50% probability, the drawn norm type is the meritocratic norm (split the total income in proportion with the (expected) performances of the two workers).[4] With 25% the drawn type follows the libertarian rule (split the total income in proprtion with the individual productions of the two workers) and with 25% an egilatarian norm is drawn (split the total income equally). We implement that meritocratic types in the Limited Information treatment use Bayesian updating[5] to infer the expected performance from the production signal and the prior distribution.

Using this setup we simulate a setting where 450 spectators make 10 distribution decisions each. 225 of them in a Full Information setting, and 225 of them in a Limited Information setting. Subsequently, a Mann-Whitney test is conducted on the per spectator average inequality level implemented over the 10 decisions. We simulated this setting for a 1000 iterations and had the Mann-Whitney test produce a p-value below 0.05 in 96% of the cases. We conclude that if the fairness decision model we use is an accurate approximation of the choices to expect in the experiment, we should expect our main nonparametric test to have a power of around 0.95.

References

[1] Almas, I., Cappelen, A.W., and Tungodden, B. (2015): "Are Americans more meritocratic and eciency-seeking than Scandinavians?." Working paper.

[2] Cappelen, A. W., Hole, A. D., Srensen, E. ., and Tungodden, B. (2005). "The pluralism of fairness ideals: An experimental approach." American Economic Review, 2007, 97(3), 818-827.

[3] Cappelen, A. W., Sørensen, E. ., and Tungodden, B. (2010). "Responsibility for what? Fairness and individual responsibility." European Economic Review, 54(3), 429-441.

[4] Cappelen, A. W., Fest, S., Sørensen, E. ., and Tungodden, B. (2016). "Choice and personal responsibility: What is a morally relevant choice?".Working paper.

[5] Dana, J., Weber, R. A., and Kuang, J. X. (2007). "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness." Economic Theory, 33(1), 67-80.

[6] Moebius, M., Niederle, M., Niehaus, P., and Rosenblat, T. (2005). "Managing Self-Condence: Theory and Experimental Evidence," mimeo.

[4] This would be approximately in line with type estimations from Cappelen et al. (2010) and Almas et al (2015).

[5] Here we do not assume any base-rate neglect.

[7] Savage, L. J. (1971). "Elicitation of personal probabilities and expectations." Journal of the American Statistical Association, 66(336), 783-801. 14

Appendix: derivation of theoretical prediction

Assume each worker's performance is drawn from a normal distribution with mean μ_p and standard deviation σ_p. Next to this, for each worker a random factor ε is drawn, individually, from a normal distribution with mean 0 and standard devation σ_ε. The performance and random factor are added to form the earnings x. This means that x will have, for each worker, a normal prior distribution with mean μ_p and standard deviation $\sigma_x = \sqrt{\sigma_p^2 + \sigma_\varepsilon^2}$ (for now we will still use σ_x). We want to determine $E(p|x)$ that is, what is the expected performance of a worker given information about the worker's earnings x.

From Bayes rule we know that the pdf of the posterior distribution should be: $f(p|x) = \frac{f(p)\cdot f(x|p)}{P(X=x)}$. Now we have $f(p) = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_p}} \cdot e^{-\frac{(p-\mu_p)^2}{2\cdot\sigma_p^2}}$ and $f(x|p) = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_\varepsilon}} \cdot e^{-\frac{(x-p)^2}{2\cdot\sigma_\varepsilon^2}}$. Let $\propto$ mean 'equal up to multiplying with a constant'. The product of the two becomes $f(p) \cdot f(x|p) = \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_p}} \cdot e^{-\frac{(p-\mu_p)^2}{2\cdot\sigma_p^2}} \cdot \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_\varepsilon}} \cdot e^{-\frac{(x-p)^2}{2\cdot\sigma_\varepsilon^2}}$

$$f(p|x) = \frac{f(p)\cdot f(x|p)}{P(X=x)} \propto f(p)\cdot f(x|p)$$

$$\begin{aligned}
f(p)\cdot f(x|p) &= \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_p}} \cdot e^{-\frac{(p-\mu_p)^2}{2\cdot\sigma_p^2}} \cdot \frac{1}{\sqrt{2\cdot\pi\cdot\sigma_\varepsilon}} \cdot e^{-\frac{(x-p)^2}{2\cdot\sigma_\varepsilon^2}} \\
&= \frac{1}{2\cdot\pi\cdot\sigma_p\cdot\sigma_\varepsilon} \cdot e^{-\left(\frac{(p-\mu_p)^2}{2\cdot\sigma_p^2}+\frac{(x-p)^2}{2\cdot\sigma_\varepsilon^2}\right)} \\
&\propto e^{-\left(\frac{(p-\mu_p)^2}{2\cdot\sigma_p^2}+\frac{(x-p)^2}{2\cdot\sigma_\varepsilon^2}\right)} \\
&= e^{-\left(\frac{\sigma_\varepsilon^2\cdot(p-\mu_p)^2+\sigma_p^2\cdot(x-p)^2}{2\cdot\sigma_p^2\cdot\sigma_\varepsilon^2}\right)} \\
&= e^{-\left(\frac{\sigma_\varepsilon^2\cdot(p-\mu_p)^2+\sigma_p^2\cdot(x-p)^2}{2\cdot\sigma_p^2\cdot\sigma_\varepsilon^2}\right)} \\
&= e^{-\left(\frac{\sigma_\varepsilon^2\cdot\left(p^2-2\cdot\mu_p\cdot p+\mu_p^2\right)+\sigma_p^2\cdot\left(p^2-2\cdot x\cdot p+x^2\right)}{2\cdot\sigma_p^2\cdot\sigma_\varepsilon^2}\right)} \\
&= e^{-\left(\frac{\left(\sigma_\varepsilon^2+\sigma_p^2\right)\cdot p^2-2\cdot\left(\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x\right)\cdot p+\sigma_\varepsilon^2\cdot\mu_p^2+\sigma_p^2\cdot x^2}{2\cdot\sigma_p^2\cdot\sigma_\varepsilon^2}\right)}
\end{aligned}$$

$$= e^{-\left(\frac{p^2-2\cdot\left(\frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x}{\sigma_\varepsilon^2+\sigma_p^2}\right)\cdot p+\frac{\sigma_\varepsilon^2\cdot\mu_p^2+\sigma_p^2\cdot x^2}{\sigma_\varepsilon^2+\sigma_p^2}}{2\cdot\frac{\sigma_p^2\cdot\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_p^2}}\right)}$$

$$\propto e^{-\frac{\left(p-\frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x}{\sigma_\varepsilon^2+\sigma_p^2}\right)^2}{2\cdot\frac{\sigma_p^2\cdot\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_p^2}}}$$

So we have that:

$$f\left(p|x\right) \propto e^{-\frac{\left(p-\frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x}{\sigma_\varepsilon^2+\sigma_p^2}\right)^2}{2\cdot\frac{\sigma_p^2\cdot\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_p^2}}}$$

Given that $f\left(p|x\right)$ is a pdf and must integrate to 1 and the right hand expression is that of a normally distributed pdf up to a multiplication with a term not containining p, we derive that $f\left(p|x\right) = \frac{1}{\sqrt{2\cdot\pi}\cdot\sigma_{p|x}} \cdot e^{-\frac{\left(p-\mu_{p|x}\right)^2}{2\cdot\sigma_{p|x}^2}}$ with $\mu_{p|x} = \frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x}{\sigma_\varepsilon^2+\sigma_p^2}$ and $\sigma_{p|x} = \sqrt{\frac{\sigma_p^2\cdot\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_p^2}}$.

Now let's assume that a meritocratic spectator will implement a level of inequality that is proportional to $|p_1 - p_2|$ a.k.a the absolute performance difference of the two workers in the full information setting, and proportional to $\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|$ in the limited information setting. So we want to compare $E\left(|p_1 - p_2|\right)$ and $E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right)$.

$$E\left(|p_1 - p_2|\right) = 2\cdot\sigma_p\cdot\sqrt{\frac{2}{\pi}} = \sigma_p\cdot\sqrt{\frac{8}{\pi}}$$

For $E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right)$ we have:

$$E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right) = E\left(\left|\frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x_1}{\sigma_\varepsilon^2+\sigma_p^2} - \frac{\sigma_\varepsilon^2\cdot\mu_p+\sigma_p^2\cdot x_2}{\sigma_\varepsilon^2+\sigma_p^2}\right|\right)$$

$$= E\left(\left|\frac{\sigma_p^2\cdot x_1 - \sigma_p^2\cdot x_2}{\sigma_\varepsilon^2+\sigma_p^2}\right|\right)$$

$$= \frac{\sigma_p^2}{\sigma_\varepsilon^2+\sigma_p^2}\cdot E\left(|x_1 - x_2|\right)$$

Now $x \sim N\left(\mu_p, \sigma_p^2 + \sigma_\varepsilon^2\right)$, which gives:

$$E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right) = \frac{\sigma_p^2}{\sigma_\varepsilon^2+\sigma_p^2}\cdot 2\cdot\sqrt{\sigma_p^2+\sigma_\varepsilon^2}\cdot\sqrt{\frac{2}{\pi}}$$

$$= \frac{\sigma_p^2}{\sqrt{\sigma_p^2 + \sigma_\varepsilon^2}} \cdot \sqrt{\frac{8}{\pi}}$$

As we should expect this expression becomes equal to $E\left(|p_1 - p_2|\right) = 2 \cdot \sigma_p \cdot \sqrt{\frac{8}{\pi}}$ if $\sigma_\varepsilon = 0$ and we are in essentially a full information situation. If $\sigma_\varepsilon > 0$ then we see that $E\left(\left|\mu_{p_1|x_1} - \mu_{p_2|x_2}\right|\right) < E\left(|p_1 - p_2|\right)$and so on average a bayesian spectator should infer smaller differences in expected performances of the two workers in a limited information setting.