# Online Appendix for "Fair Shares and Selective Attention"

Dianna R. Amasino, Davide D. Pace, Joël J. van der Weele

**Table of contents**

# A Model

This section develops a model that investigates the relationship between attention and redistribution decisions and between selective attention and self-serving biases. The proofs of the results are provided in Appendix 3..

## 1. Set-up

Imagine two agents, a dictator (she) and a recipient (he), indicated with subscript $i = 1$ and $i = 2$ respectively. In the production phase, the agents produce $w_i$ leading to joint endowment $W = w_1 + w_2$. We model the role of effort and luck multiplicatively, as in the experiment: $w_i = e_i l_i$. Here, $e_i \in [\underline{e}; \bar{e}]$ is the effort exerted by $i$, where $\underline{e} > 0$ is the minimum possible effort and $\bar{e} > \underline{e}$ is the highest possible effort. The luck component is the multiplier on effort $l_i \in L, H$ with $0 < L < H$. The dictator and the recipient differ in their luck, i.e. $l_1 \neq l_2$. We call the dictators for which $l_1 = H$ "Advantaged" and the dictators for which $l_1 = L$ "Disadvantaged". After the production phase, the dictator decides the allocation of the endowment $(x_1, x_2)$ such that $x_1 + x_2 = W$.

The dictator knows $W$ and whether she is Advantaged or Disadvantaged, that is, whether $l_1 > l_2$ or vice versa. She does not know the exact values of $l_1$ and $l_2$, but she knows they have a distribution $f_{l_1}(l_1)$ and $f_{l_2}(l_2)$. Moreover she believes that $e_i \sim f_e(e_i)$. The effort levels of the two agents are independent and identically distributed. To resolve the uncertainty about $e_i$ and $l_i$ the dictator can access a signal about $(e_1, e_2)$ and another signal about $(w_1, w_2)$. These signals resolve uncertainty fully and are available for free, although paying attention to them may be costly, as we discuss below.[23]

**Dictator preferences.** The dictator's preferences are represented by

$$U(x_1, \mathbf{t}) = u(x_1) - g(x_1 - r(\boldsymbol{t})) - C(\mathbf{t}, \bar{\mathbf{t}}). \tag{A.1}$$

Here, $u(x_1)$ is the utility from her monetary allocation, which is increasing and concave. The second component captures guilt from an unfair allocation, which may depend on attention. The last component is an attentional cost. We now discuss these two last components in detail.

[23] This setup follows our experimental design. Our model also works if dictators are perfectly informed, and attention to different kinds of information only serves to contemplate the associated fairness criterion.

**Fairness.** The term $g(x_1 - r(\boldsymbol{t}))$ indicates the guilt cost the dictator pays if she keeps more than $r \in [0; w]$, the amount the dictator considers to be fair to keep for herself. We assume that $g : [0; w] \to \mathbb{R}_+$, $g(x_1 - r) = 0$ if $x_1 \leq r$, and $g(x_1 - r) > 0$ if $x_1 > r$. $g(x_1 - r)$ is twice differentiable, increasing and strictly convex if $x_1 > r$. Modeling fairness concerns as disutility (guilt) from the difference between the actual and the fair share is common in the literature (Konow, 2000; Rodriguez-Lara and Moreno-Garrido, 2012; Cappelen et al., 2007, 2013).

We assume the fair amount depends on a weighted sum of three different fairness criteria:

$$r(\mathbf{t}) = \pi_\mu(\mathbf{t})x_\mu + \pi_\lambda(\mathbf{t})x_\lambda + \pi_\eta(\mathbf{t})x_\eta.$$

Here, $x_k$ is fair amount the dictator should keep according to criterion $k \in \{\mu, \lambda, \eta\}$:

- The **meritocratic criterion** $x_\mu := W\frac{e_1}{e_1+e_2}$ prescribes keeping an amount proportional to the dictator's effort.
- The **libertarian criterion** $x_\lambda := w_1$ prescribes keeping an amount proportional to the dictator's output, without correcting for luck.
- The **egalitarian criterion** $x_\eta := \frac{W}{2}$ prescribes keeping half of the output.

The weight $\pi_k(\mathbf{t})$ depends on attention vector $\boldsymbol{t} = \{t_\mu, t_\lambda, t_\eta\}$, where $t_k$ indicates the timespan that dictator attends to criterion $k$ (see more details below). We assume $\pi_k \geq 0$ and that $\frac{\partial \pi_k}{\partial t_k} > 0$, so the weights increase in the attention paid to the corresponding criterion. We normalize the vector $\boldsymbol{\pi}$ such that $\sum_{k\in K} \pi_k(\mathbf{t}) = 1 \; \forall \; \boldsymbol{t}$, this implies that $\frac{\partial \pi_k}{\partial t_{-k}} < 0$.

This way of modelling distortions due to attention is adapted from Bordalo, Gennaioli and Shleifer (2021). The positive relation between attention and decision weights is supported by Pärnamets et al. (2015), who exogenously manipulates the time participants spend looking at two statements regarding controversial moral topics. They find that participants are more likely to endorse the statement that they look at longer. Ghaffari and Fiedler (2018) replicated and extended this finding. Section B. discusses how visual attention changes what the dictators consider a fair allocation in our experiment.

**Attention.** The attention vector $\boldsymbol{t} = \{t_\mu, t_\lambda, t_\eta\}$ captures two types of attention. First, it captures visual inspection of information. Thus, we will denote by $t_\mu$ the time spent accessing information about efforts $e_i$, as this is relevant exclusively for the meritocratic criterion. Similarly, we denote by $t_\lambda$ the time spent attending to outputs $w_i$, as this is relevant exclusively for

the libertarian criterion. Second, visual information may be accompanied with various types of information processing and introspective contemplation to evaluate the proper use of the criterion. Indeed, the well-established eye-mind theory shows that visual attention is accompanied by the processing of the underlying information (Just and Carpenter, 1980). We assume all these aspects are captured by $t$. This also allows us to model attention egalitarian split $t_\eta$, which does not require visual attention to any of the production data, but is relevant for Proposition 3 below. Proposition 1, 2, and 4 go through when we drop attention to the egalitarian criterion and we assume that there is a fixed weight the dictator gives to the egalitarian criterion.

We make the following assumptions about $t$.

1. **Attention budget.** The dictator has a total time of $T$ to attend to information, and needs to spend all this time looking at information such that $t_\mu + t_\lambda + t_\eta = T$.

2. **Top-down control.** The dictator has control over her attention. That is, she chooses a vector of attention $\boldsymbol{t} \in S$ , where $S$ is a 2-simplex of edge length T.

3. **Bottom-up salience.** Attentional control is costly, as certain states may be salient and naturally attract attention. The importance of salience for decision making are documented in an well established literature in psychology and a growing literature in economics Bordalo, Gennaioli and Shleifer (2021). To capture this, we assume there exists a *default* bottom-up attention pattern $\bar{\boldsymbol{t}} = \{\bar{t}_\mu, \bar{t}_\lambda, \bar{t}_\eta\}$ with $\bar{t}_\lambda > 0$, $\bar{t}_\mu > 0$, and $\bar{t}_\eta \geq 0$.[24] When the dictator deviates from default $\bar{\boldsymbol{t}}$, she pays a cost $C(\boldsymbol{t}, \bar{\boldsymbol{t}}) : S \to \mathbb{R}$, which is twice differentiable and increasing in $|t_k - \bar{t}_k| \ \forall \ k \in K$. This cost captures both the concentration costs of manipulating attention and the psychological cost of diverting attention in order to self-deceive about the size of $r$ and hence increase $x_1$. We normalize $C(\bar{\boldsymbol{t}}, \bar{\boldsymbol{t}}) = 0$.

4. **Curiosity.** $C(\boldsymbol{t}, \bar{\boldsymbol{t}}) > u(W)$ if $t_\lambda = 0$ or if $t_\mu = 0$. This assumption models curiosity, as it assures ignorance is too costly for the dictator. Golman et al. (2021) shows how curiosity is an important driver for information acquisition, in particular when it is salient that information is available. In our data, most dictators access all information (see Section E. for more details).

[24]Note that default $\bar{\boldsymbol{t}}$ is a function of the information. That is: $\bar{\boldsymbol{t}}(e_1, e_2, w_1, w_2) : \mathbb{R}^4 \to S$. Hence $\bar{\boldsymbol{t}}$ depends only on the decision making environment, and not on the dictators characteristics (e.g. whether she is Advantaged or Disadvantaged). That is $\bar{\boldsymbol{t}}(e_1, e_2, e_1H, e_2L) = \bar{\boldsymbol{t}}(e_1, e_2, e_1L, e_2H)$. Moreover, the dependence of the bottom-up vector of attention on the information highlights that the bottom-up process influence the dictator only if she access the information. We don't formally model what happens when the dictator avoids all or part of the information because Lemma 1 shows that the dictator always access the information about $(e_1, e_2, w_1, w_2)$.

5. **Speed of learning.** Minimal attention is needed to acquire the information about $(w_1, w_2)$, and $(e_1, e_2)$ and resolve all the uncertainty. If $t_\lambda > 0$, the dictator knows the exact values of $(w_1, w_2)$. If $t_\mu > 0$, the dictator knows the exact value of $(e_1, e_2)$. This assumption reflects the fact that the information is very simple. Four numbers are all that the dictators have to learn. In the experiment, these numbers are mostly one or two digits.

**Timeline.** The timeline is as follows:

$\tau = 0$ Production task.

$\tau = 1$ The dictator receives perfect information on $W$ and on whether she is Advantaged or Disadvantaged. Furthermore, she can allocate her attention, and access information about $e_1$, $e_2$, $w_1$, and $w_2$. The time she spends on the different types of information is captured by $\mathbf{t}$.

$\tau = 2$ The dictator splits $W$ in $x_1$ and $x_2$.

The dictator maximizes her utility by choosing $\mathbf{t}$ and $x_1$ sequentially. To solve the model, we therefore work backwards, first computing the optimal choice for a given level of attention, and then maximizing the level of attention given the resulting choice.

## 2. Results.

We first show that our model predicts selective attention and that selective attention changes allocation decisions.. Then we we turn to the impact of implementing exogenous restrictions on attention, as in our constrained focus treatments. Finally we demonstrate that Advantaged dictators keep a larger amount for themselves than the Disadvantaged ones

**Selective attention.** Let's call $t_\mu^{*A}$ and $t_\lambda^{*A}$ the optimal level of attention to information about merit and about outcome if the dictator is Advantaged and $t_\mu^{*D}$ and $t_\lambda^{*D}$ the optimal level of attention if she is Disadvantaged. We can define $\Delta$ Attention$^A = t_\mu^{*A} - t_\lambda^{*A}$ and $\Delta$ Attention$^D = t_\mu^{*D} - t_\lambda^{*D}$.

**Proposition 1** (Selective Attention)**.** $\Delta$ *Attention*$^A$ $<$ $\Delta$ *Attention*$^D$*. That is, compared to Disadvantaged dictators, Advantaged ones spend relatively less time looking at information about effort and relatively more time looking at information about outcome.*

Intuitively, dictators distort their attention to believe that they deserve a larger share of the endowment and hence reduce their guilt over keeping a larger share. Advantaged and Disadvantaged dictators, however, distort attention in opposite directions. The Advantaged dictators move attention from merit information to outcome information because they receive more if they implement a libertarian rather than a meritocratic split. The opposite is true for the Disadvantaged dictators: they shift their attention from the outcome to the merit information because they receive more from a meritocratic rather than from a libertarian split.

**Corollary 1.1** (Attention as a mediator of self-serving biases)**.** *Selective attention allows dictators to act more selfishly.*

This result follows immediately from Proposition 1. Selective attention reduces the marginal guilt cost for any amount the dictator keeps for herself. As a result, the dictator keeps more.

**Restricting attention.** Our objective in this paragraph is to check whether the model predicts that Advantaged agents receive more money in the Outcome Focus treatment than in the Merit Focus treatment. To do so we need to formalize our two attention manipulations in our experiment, let's call them *Mer* and *Lib*. Without loss of generality, we assume that *Mer* is the manipulation that restricts attention to $(e_1, e_2)$, while *Lib* restricts attention to $(w_1, w_2)$. Hence *Mer* models the Outcome Focus treatment, and *Lib* models the Merit Focus treatment. The manipulations restrict the set of vectors of attention among which the dictators can choose. Let's call $\hat{S}^{Mer} \subset S$ and $\hat{S}^{Lib} \subset S$ the two sets of feasible vectors of attention when the manipulations are in place. Moreover, let's define $t_\mu^{*Mer}$ and $t_\lambda^{*Mer}$ the optimal attention to the meritocratic and libertarian criteria in $\hat{S}^{Mer}$. Similarly, define $t_\mu^{*Lib}$ and $t_\lambda^{*Lib}$ the optimal attention to the meritocratic and libertarian criteria in $\hat{S}^{Lib}$. Finally, define $\Delta\text{Attention}^{Mer} = t_\mu^{*Mer} - t_\lambda^{*Mer}$ and $\Delta\text{Attention}^{Lib} = t_\mu^{*Lib} - t_\lambda^{*Lib}$.

We can now go back to our experiment and study three empirical properties of the attention manipulations, to put some reasonable assumptions on the effect of the manipulation. First, from Table 2, we can see that $\Delta\text{Attention}^{Lib} > \Delta\text{Attention}^{Mer}$. Second, from the same table, we see that the dictators spend a similar amount of time looking at Merit and Outcome information in the Merit Focus and in the Outcome Focus treatments. We can approximate this finding assuming that $t_\mu^{*Mer} + t_\lambda^{*Mer} = t_\mu^{*Lib} + t_\lambda^{*Lib}$. Finally, from the design we derive that the total time people spend thinking about the different criteria is the same independently of the treatment: subjects have 6 seconds on the information screen and then they are automatically redirected to the decision screen. Hence, $\sum_{k \in K} t_k^{*Mer} = \sum_{k \in K} t_k^{*Lib}$.

The proposition below shows that these three properties of the attention manipulations are sufficient conditions for Advantaged agents to receive more money under attention manipulation *Mer* than under attention manipulation *Lib*.

**Proposition 2** (Effect of constrained Attention on Allocation)**.** *If* $\Delta Attention^{Lib} > \Delta Attention^{Mer}$, $t_\mu^{*Lib} + t_\lambda^{*Lib} = t_\mu^{*Mer} + t_\lambda^{*Mer}$, *and* $\sum_{k \in K} t_k^{*Lib} = \sum_{k \in K} t_k^{*Mer}$, *then the Advantaged agents receives more money if* $\boldsymbol{t} \in \hat{S}^{Mer}$ *than if* $\boldsymbol{t} \in \hat{S}^{Lib}$.

The result obtains because a lower $\Delta$Attention decreases the weight the dictator gives to the meritocratic criterion and increases the weight she gives to the libertarian one. As a result, Advantaged dictators keep more money and Disadvantaged ones keep less.

While Proposition 2 predicts our main empirical result, a puzzling finding from our experiment is that the Advantaged dictators react more to our attention manipulations than Disadvantaged dictators. The proposition below shows that the model predicts this finding under some reasonable simplifying assumptions about a) the functional form of the utility function b) the attention process c) the characteristics of the attention manipulations. These assumptions are sufficient but not necessary for deriving the result.

**Assumption 1** (Simplifying assumptions about the utility function)**.**

$$
\begin{aligned}
&u(x_1) = x_1 \\
&g(x_1 - r) = \frac{1}{2}\beta g(x_1 - r)^2 \\
&C(\boldsymbol{t}, \bar{\boldsymbol{t}}) = \frac{1}{2}\gamma[(t_\lambda - \bar{t_\lambda})^2 + (t_\mu - \bar{t_\mu})^2 + (t_\eta - \bar{t_\eta})^2]
\end{aligned}
$$

The first line of the assumption states utility function is linear in money. This is a good approximation for small stakes like the ones in our experiment (Rabin, 2000). The second line says that the guilt function is quadratic. This is a common assumption in the literature on fairness norms (Cappelen et al., 2007; Bortolotti et al., 2017). The third line states that the cost of attention distortion is a sum of quadratic costs. We chose this quadratic form for consistency with the functional form of the guilt function.

**Assumption 2** (Simplifying assumptions about the attention process)**.**

$$r(\boldsymbol{t}) = \frac{t_\lambda}{t_\lambda + t_\mu + t_\eta} x_\lambda + \frac{t_\mu}{t_\lambda + t_\mu + t_\eta} x_\mu + \frac{t_\eta}{t_\lambda + t_\mu + t_\eta} x_\eta$$
$$\bar{\boldsymbol{t}} = \{\frac{T}{3}, \frac{T}{3}, \frac{T}{3}\}$$

The first line assumes that the weight given to each criterion is proportional to the time the dictator spends on it. The second that the bottom up processes are such that dictators allocate equal time to all criteria if they don't distort their attention. The key feature of these assumptions is that they treat the three criteria in the same way and assure that our result is not due to us considering one of the criteria as special.

**Assumption 3** (Simplifying assumptions about the attention manipulation)**.** *The attention manipulation restricts the attention to the information relevant for criterion* $k \in \{w, \eta, g\}$ *to* $t_k = \hat{t}$*. In addition:*

- *The attention manipulation is always binding*
- *The attention the dictator is forced to divert from criterion* $k$*, that is* $t_k - \hat{t}$*, is equally split among the other two criteria.*

This assumption introduces an attention manipulation similar to the one we used in the experiment but simpler to analyse. In the experiment, we introduced a tighter constraint on one criterion and a softer one on another criterion. Here, for simplicity, we assume that the manipulation only constrains one criterion. Moreover, the first bullet point excludes cases in which the manipulation is not binding. The second bullet point assumes that the dictators are forced to equally split among the other two criteria the attention that they have to redirect. This mechanical redirection of attention simplifies our analysis considerably because spares us from analysing how the dictators re-optimize their attention under the attention restriction. Yet, the resulting vector of attention is likely close to what we would have obtained without this assumption. $C(\boldsymbol{t}, \bar{\boldsymbol{t}})$ is increasing and convex pushing the dictators to reallocate the attention more or less evenly across the two other criteria.

**Assumption 4.** *[Interior solution] Both with and without the manipulation, dictators keep less than the entire pie.*

This assumption reflects dictators' behavior in our experiment: the dictator keeps the entire pie in only 2.8% of the trials.

To state the next proposition we need some additional notation. Let's define $x_1^{A-Mer}$ the amount the Advantage dictator keeps if her attention to the merit information is restricted to $\hat{t}$. Instead, let's define $x_1^{A-Lib}$ the share she keeps if we restrict to $\hat{t}$ her attention to the outcome information. $x_1^{D-Mer}$ and $x_1^{D-Lib}$ indicate the behavior of the Disadvantaged dictator under the same restrictions. The combined effect of the attention manipulations for an Advantaged dictator is then given by $|x_1^{A-Mer} - x_1^{A-Lib}|$. For a Disadvantaged it is $|x_1^{D-Mer} - x_1^{D-Lib}|$.

**Proposition 3** (The differential effect of the attention manipulation on Advantaged and Disadvantaged dictators)**.** *In Expectation, under Assumptions 1, 2, 3 and 4, the combined effect of the attention manipulations is larger for the Advantaged dictators. That is,* $E[|x_1^{A-Mer} - x_1^{A-Lib}|] > E[|x_1^{D-Mer} - x_1^{D-Lib}|] > 0$.

This result obtains because in most cases the Advantaged dictators have a larger incentive to distort their attention towards the libertarian criterion than the Disadvantaged dictators to distort their attention towards the meritocratic criterion. Hence the behavioral effect of restricting attention towards the libertarian criterion for the Advantaged dictators is larger than the behavioral effect of restricting attention towards the meritocratic criterion for the Disadvantaged dictators. Moreover, for those same cases, it is optimal for the Advantaged dictators to distort their attention *away* from the egalitarian criterion, while it is optimal for the Disadvantaged dictators to distort their attention *towards* the egalitarian criterion. As a consequence, the Advantaged dictators spend more time on the meritocratic and libertarian criteria which are the ones affected by the attention manipulation. More time on these criteria implies that the manipulation shifts the attention and, hence, the behavior of the Advantaged dictators to a larger extent.

**The effect of status of behavior.** One last thing that we want to show is that, in expectation, the Advantaged dictators keep a larger amount of the pie than the Disadvantaged dictators do. This is the patterns established in (Konow, 2000) and reproduced in follow-up papers.

To simplify the proof, we derive this result under the same assumptions we used for Proposition 3.

**Proposition 4.** *Under Assumptions 1, 2, and 4, in expectation the Advantaged dictators keep a larger amount than the Disadvantaged dictators. That is* $E(x_1^{*A}) > E(x_1^{*D})$

The Advantaged dictators keep a larger amount of the pie because there is a criterion (the libertarian one) that favors them in expectation. Hence, turning their attention to this criterion raises the share they consider fair to keep above 50 percent. By contrast, the Disadvantaged dictators most favorable criteria are the egalitarian or meritocratic ones, but these do not lead to a bigger than 50/50 split in expectation.

Thus, our model predicts the well-documented fact that status affects the way people redistribute money. Moreover, because the mechanism operates through shifts in subjective fairness criteria, the model could be extended to predict that the effect carries over to any subsequent impartial decisions. Note that (Konow, 2000) already modeled this finding. Our contribution is to ground this effect in an attentional mechanism.

## 3. Proofs

### 3..1 Lemmas

Before proving Propositions 1 and 2, it is useful to prove some lemmas that will come in handy for the following derivations.

**Lemma 1** (Perfect information)**.** *Dictators always spend a positive amount of time looking at information relevant to the libertarian and meritocratic criteria. Hence, know the exact value of* $w_1$*,* $w_2$*,* $e_1$*, and* $e_2$*.*

*Proof.* By assumption, we know that $C(\tilde{\boldsymbol{t}}) > u(W)$ where $\tilde{\boldsymbol{t}} \ : \ t_\lambda = 0 \vee t_\mu = 0$. This means that $U(x_1, \tilde{\boldsymbol{t}}) < 0 \ \forall \ x_1 \in [0, W]$. However, the dictator can get a higher payoff if she doesn't distort her attention. In fact, $U(x_1, \bar{\boldsymbol{t}}) = 0$. Hence the optimal vector of attention $\boldsymbol{t^*}$ is such that $t_\lambda > 0$ and $t_\mu > 0$.

By definition, $t_\lambda$ is the time the dictators spend on the information about $w_1$ and $w_2$, while $t_\mu$ is the time they spend on the information about $e_1$ and $e_2$. By assumption, spending any positive time is enough to acquire the information. Hence, the dictators know the exact values of $w_1$, $w_2$, $e_1$, and $e_2$.

□

Intuitively, it is optimum for the dictator to access the perfectly informative signals about effort and output because not doing so would have an extremely high cost.

**Lemma 2** (The effect of attention on $r(\boldsymbol{t})$)**.** *If the fairness criteria* $k \in \{\lambda, \mu, \eta\}$ *are ordered such that* $k_1 > k_2 > k_3$*: Then:* $\dfrac{\partial r(\boldsymbol{t})}{\partial t_j} - \dfrac{\partial r(\boldsymbol{t})}{\partial t_z} > 0$ *if* $j > z$

*Proof.* Take $j > z$ and $\boldsymbol{t}, \boldsymbol{t'} \in S \ : \ \boldsymbol{t} = \boldsymbol{t'} + \boldsymbol{\xi}$. Where $\boldsymbol{\xi} \ : \ \xi_j = \varepsilon, \xi_z = -\varepsilon, \xi_{k \neq j,z} = 0$ with $\varepsilon > 0$. By assumption, we know that $\pi_k(\boldsymbol{t})$ is increasing in $t_k$ and decreasing in $t_{-k}$. Hence:

$$\begin{aligned}
&\pi_j(\boldsymbol{t}) > \pi_j(\boldsymbol{t}) \\
&\pi_z(\boldsymbol{t}) < \pi_j(\boldsymbol{t}) \\
&\pi_k(\boldsymbol{t}) = \pi_k(\boldsymbol{t}) \ \wedge \ k \neq j, z
\end{aligned}$$

We know that $r(\boldsymbol{t}) = \sum_{k \in K} \pi_k k_1$, and that $j_1 > z_1$ so it must be that $r(\boldsymbol{t}) - r(\boldsymbol{t'}) > 0$. As the last inequality must hold for every $\varepsilon > 0$ including for $\varepsilon$ infinitesimally small, it follows that:

$$\frac{\partial r(\boldsymbol{t})}{\partial t_j} - \frac{\partial r(\boldsymbol{t})}{\partial t_z} > 0 \tag{A.2}$$

□

Intuitively, if the dictator shift attention from one criterion that posit she should keep less to one that posits she should keep more, she increases the weight she gives to the latter. As such the amount she considers fair to keep increases. This intuition holds also for infinitesimal amounts.

**Corollary 2.1.** *For Advantaged* $\dfrac{\partial r(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial r(\boldsymbol{t})}{\partial t_\mu} > 0$*, for disadvantaged* $\dfrac{\partial r(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial r(\boldsymbol{t})}{\partial t_\mu} < 0$

*Proof.*

$$\begin{aligned}
&sign\left[x_\lambda - x_\mu\right] = \\
=&sign\left[e_1 l_1 - \frac{(e_1 l_1 + e_2 l_2) e_1}{e_1 + e_2}\right] \\
=&sign\left[e_1 e_2 (l_1 - l_2)\right]
\end{aligned}$$

Since $e_1, e_2 > 0$ by assumption, the sign is positive if $l_1 > l_2$ and negative if $l_1 < l_2$ By definition Advantaged dictators are those for whom $l_1 > l_2$ and, vice versa, Disadvantaged dictators are those for whom $l_1 > l_2$. Hence, by Lemma 2 we conclude that for Advantaged dictators $\dfrac{\partial r(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial r(\boldsymbol{t})}{\partial t_\mu} > 0$, while for Disadvantaged ones $\dfrac{\partial r(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial r(\boldsymbol{t})}{\partial t_\mu} < 0$. □

**Lemma 3** (Keeping more than the fair share)**.** *In the optimum, the dictator keeps more than the share she considers fair. That is* $x_1^*(\boldsymbol{t^*}) > r(\boldsymbol{t^*})$*.*

*Proof.* The dictator chooses $\boldsymbol{t}$ and $x_1$ sequentially. So at the moment of choosing $x_1$, $\boldsymbol{t}$ is fixed and the cost of attention are sunk. As such the maximization problem for $x_1$ is the following.

$$x_1^*(\boldsymbol{t}) = \arg\max_{x_1 \in [0;w]} [u(x_1) - g(x_1 - r)]$$

As the maximand is defined on a closed and bounded interval, the Weierstrass theorem assures the existence of a solution. To characterize the solution for this problem let's take the first derivative of the maximand:

$$\frac{\partial u(x_1)}{\partial x_1} = \frac{\partial g(x_1 - r)}{\partial x_1} \tag{A.3}$$

The LHS indicates the marginal benefit of increasing $x_1$, this benefit is strictly positive because $u(x_1)$ is increasing. The RHS, instead, represent the marginal cost of increasing $x_1$ due to guilt. $g(x_1 - r) = 0$ if $x_1 \leq r(\boldsymbol{t})$ and positive and increasing elsewhere. Hence the marginal cost is equal to 0 if $x_1 \leq r(\boldsymbol{t})$ while it is positive if $r(\boldsymbol{t}) > x_1$.

Since $g(x_1 - r)$ is twice differentiable, it follows that it's first derivative is continuous. Hence:

$$\lim_{x_1 \to r(\boldsymbol{t})^-} \frac{\partial g(x_1 - r)}{\partial x_1} = \frac{\partial g(r - r)}{\partial x_1} = \lim_{x_1 \to r(\boldsymbol{t})^+} \frac{\partial g(x_1 - r)}{\partial x_1} = 0$$

Because $\lim_{x_1 \to r(\boldsymbol{t})^-} \frac{\partial g(x_1 - r)}{\partial x_1} = 0$. As a consequence $x_1^*(\boldsymbol{t}) > r(\boldsymbol{t})$. □

The lemma derives from the fact that keeping a bit more than what the dictator think is fair generates first order gains but only infinitesimal costs.

**Corollary 3.1.** *The dictator keeps strictly more than zero. That is* $x_1 > 0$

*Proof.* $r(\boldsymbol{t}) \geq 0$ because $r(\boldsymbol{t})$ is a linear combination of positive numbers. Hence $x_1 > r(\boldsymbol{t}) \geq 0$. □

**Lemma 4** (The relationship between $r$ and $x_1$)**.** *If* $x_1^* < W$*, then* $\dfrac{\partial x_1^*}{\partial r} > 0$*. That is, unless the dictator is already keeping everything, the share she keeps is increasing in the share she considers fair to keep.*

*Proof.* If we assume that $x_1^* < W$, then by Corollary 3.1, $0 < x_1 < W$. Hence, Equation A.3 above gives us the necessary and sufficient condition for $x_1^*$. The sufficiency of the condition follows from the concavity of $u(x_1) - g(x_1 - r)$. In fact, $u(x_1)$ is concave and $g(x_1 - r)$ is convex by assumption.

We can now differentiate Equation A.3 w.r.t. $r$ and obtain:

$$\frac{\partial^2 u(x_1)}{\partial^2 x_1}\frac{\partial x_1}{\partial r} = \frac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r)}\left[\frac{\partial x_1}{\partial r} - 1\right] \tag{A.4}$$

$$\frac{\partial x_1}{\partial r} = \frac{-\dfrac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r)}}{\dfrac{\partial^2 u(x_1)}{\partial^2 x_1} - \dfrac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r)}} \tag{A.5}$$

$$\frac{\partial x_1}{\partial r} > 0 \tag{A.6}$$

Where the last step comes from the fact that $u(x_1)$ is concave and, by Lemma 3, $g(x_1 - r)$ is strictly convex if $x_1 > r$. □

**Lemma 5.** *There is an optimum* $\boldsymbol{t}^*$ *and* $\boldsymbol{t}^* \neq \bar{\boldsymbol{t}}$.

*Proof.* The maximization problem for $\boldsymbol{t}$ is given by:

$$\boldsymbol{t}^* = \arg\max_{\boldsymbol{t} \in S} U(x_1^*(\boldsymbol{t}), \boldsymbol{t}) = \arg\max_{\boldsymbol{t} \in S} u(x_1^*(\boldsymbol{t})) - g(x_1^*(\boldsymbol{t}) - \hat{r}_1(\boldsymbol{t})) - C(\boldsymbol{t}, \bar{\boldsymbol{t}})$$

$$\text{Subject to: } t_\lambda + t_\mu + t_\eta = T$$

As $S$ is a closed and bounded subset of the domain of $U(x_1^*(\boldsymbol{t}), \boldsymbol{t})$, we know that there must exist an optimum $\boldsymbol{t}^* \in S$ by the Weierstrass theorem.

The FOCs of the problem are given by

$$\begin{cases} \dfrac{\partial u(x_1^*(\boldsymbol{t}))}{\partial x_1}\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\mu} - \dfrac{\partial g(x_1^*(\boldsymbol{t}) - r(\boldsymbol{t}))}{\partial (x_1 - r)}\left[\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\mu} - \dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\mu}\right] - \dfrac{\partial C(\boldsymbol{t})}{\partial t_\mu} = \nu \\ \dfrac{\partial u(x_1^*(\boldsymbol{t}))}{\partial x_1}\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial g(x_1^*(\boldsymbol{t}) - r(\boldsymbol{t}))}{\partial (x_1 - r)}\left[\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\lambda}\right] - \dfrac{\partial C(\boldsymbol{t})}{\partial t_\lambda} = \nu \\ \dfrac{\partial u(x_1^*(\boldsymbol{t}))}{\partial x_1}\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\eta} - \dfrac{\partial g(x_1^*(\boldsymbol{t}) - r(\boldsymbol{t}))}{\partial (x_1 - r)}\left[\dfrac{\partial x_1(\boldsymbol{t})}{\partial t_\eta} - \dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\eta}\right] - \dfrac{\partial C(\boldsymbol{t})}{\partial t_\eta} = \nu \end{cases}$$

Where $\nu$ is the Lagrangian multiplier. If $\boldsymbol{t}^* = \bar{\boldsymbol{t}}$, the FOC above must hold in $\boldsymbol{t} = \bar{\boldsymbol{t}}$. Applying

the envelop theorem and substituiting $\nu$, we can rewrite one of the FOCs in $\boldsymbol{t} = \bar{\boldsymbol{t}}$ as:

$$\frac{\partial g(x_1^*(\bar{\boldsymbol{t}}) - r(\bar{\boldsymbol{t}})}{\partial(x_1 - r(\boldsymbol{t}))}\left[\frac{\partial r(\bar{\boldsymbol{t}})}{\partial t_\mu} - \frac{\partial r(\bar{\boldsymbol{t}})}{\partial t_\lambda}\right] - \frac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\mu} + \frac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\lambda} =$$

$$\frac{\partial g(x_1^*(\bar{\boldsymbol{t}}) - r(\bar{\boldsymbol{t}})}{\partial(x_1 - r(\boldsymbol{t}))}\left[\frac{\partial r(\bar{\boldsymbol{t}})}{\partial t_\mu} - \frac{\partial r(\bar{\boldsymbol{t}})}{\partial t_\lambda}\right] \neq 0$$

The first step comes from the fact that $\dfrac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\mu} = \dfrac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\lambda} = 0$. In fact, by assumption we know that $C(\boldsymbol{t}, \bar{\boldsymbol{t}})$ is increasing in $\|\boldsymbol{t} - \bar{\boldsymbol{t}}\|$ and that $C(\boldsymbol{t}, \bar{\boldsymbol{t}})$ is differentiable. Hence, $C(\boldsymbol{t}, \bar{\boldsymbol{t}})$ has a minimum in $\bar{\boldsymbol{t}}$ and in that point the $\dfrac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\mu} = \dfrac{\partial C(\bar{\boldsymbol{t}})}{\partial t_\lambda} = 0$. The last line is different from zero because $\dfrac{\partial g(x_1^*(\boldsymbol{t}) - r(\boldsymbol{t}))}{\partial(x_1 - r)} > 0$ by Lemma 3, and $\dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\mu} - \dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\lambda} \neq 0$ by Corollary 2.1. Hence we conclude that $\boldsymbol{t}^* \neq \bar{\boldsymbol{t}}$. □

We are now ready to prove Proposition 1

### 3..2 Proof of Proposition 1

*Proof.* Define as $\boldsymbol{t^{*A}}$ and $\boldsymbol{t^{*D}}$ the optimal vector of attention for the Advantaged and the Disadvantaged dictators respectively. We want to prove that $t_\mu^{*A} - t_\lambda^{*A} < t_\mu^{*D} - t_\lambda^{*D}$. To do so, we will prove that $t_\mu^{*A} - t_\lambda^{*A} < \bar{t}_\mu - \bar{t}_\lambda < t_\mu^{*D} - t_\lambda^{*D}$. Where $\bar{t}_\mu$ and $\bar{t}_\lambda$ are the same for Advantaged and Disadvantaged dictators because, by assumption, the bottom up vector of attention $\bar{\boldsymbol{t}}$ does not depend on whether the dictator is Advantaged or Disadvantaged.

Let's first prove that $t_\mu^{*A} - t_\lambda^{*A} < \bar{t}_\mu - \bar{t}_\lambda$. A sufficient condition for this inequality to hold is:

$$(t_\mu^{*A} - \bar{t}_\mu < 0 \wedge t_\lambda^{*A} - \bar{t}_\lambda \geq 0) \vee (t_\mu^{*A} - \bar{t}_\mu \leq 0 \wedge t_\lambda^{*A} - \bar{t}_\lambda > 0) \tag{A.7}$$

To prove that expression A.7 is true we will begin showing that any $\boldsymbol{t} \ : \ t_\mu > \bar{t}_\mu, \ t_\lambda < \bar{t}_\lambda$ cannot be optimum. To do so, lets define $\boldsymbol{t}, \boldsymbol{t'}, \boldsymbol{t''} \in S \ : \ t_\mu > \bar{t}_\mu, \ t_\lambda < \bar{t}_\lambda, \ \boldsymbol{t'} = \boldsymbol{t} + \boldsymbol{\xi}, \ \boldsymbol{t''} = \boldsymbol{t} + \boldsymbol{\xi} + \boldsymbol{\xi'}$. Where $\boldsymbol{\xi} \ : \ \xi_\lambda = 0, \ \xi_\mu = -\varepsilon, \ \xi_\eta = 0$ and $\boldsymbol{\xi'} \ : \ \xi'_\lambda = \varepsilon, \ \xi'_\mu = 0, \ \xi'_\eta = 0$ with $\varepsilon > 0$.

$$sign\left[U(x_1^*(\boldsymbol{t}),\boldsymbol{t}) - U(x_1^*(\boldsymbol{t''}),\boldsymbol{t''})\right] =$$

$$sign\left[U(x_1^*(\boldsymbol{t}),\boldsymbol{t}) - U(x_1^*(\boldsymbol{t'}),\boldsymbol{t'}) + U(x_1^*(\boldsymbol{t'}),\boldsymbol{t'}) - U(x_1^*(\boldsymbol{t''}),\boldsymbol{t''})\right] =$$

$$= sign\left[\lim_{\varepsilon\to 0}\left(\frac{U(x_1^*(\boldsymbol{t}),\boldsymbol{t}) - U(x_1^*(\boldsymbol{t}+\boldsymbol{\xi}),\boldsymbol{t}+\boldsymbol{\xi})}{\varepsilon}\right) + \lim_{\varepsilon\to 0}\left(\frac{U(x_1^*(\boldsymbol{t}+\boldsymbol{\xi}),\boldsymbol{t}+\boldsymbol{\xi}) - U(x_1^*(\boldsymbol{t}+\boldsymbol{\xi}+\boldsymbol{\xi'}),\boldsymbol{t}+\boldsymbol{\xi}+\boldsymbol{\xi'})}{\varepsilon}\right)\right] =$$

$$= sign\left[\frac{\partial U(x_1^*(t))}{\partial t_\lambda} - \frac{\partial U(x_1^*(t))}{\partial t_\mu}\right] =$$

$$= sign\left[\frac{\partial g(x_1-\hat{r}_1)}{\partial(x_1-\hat{r}_1)}\frac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\lambda} - \frac{\partial C(\boldsymbol{t})}{\partial t_\lambda} - \frac{\partial g(x_1-\hat{r}_1)}{\partial(x_1-\hat{r}_1)}\frac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\mu} + \frac{\partial C(\boldsymbol{t})}{\partial t_\mu}\right] =$$

$$= sign\left[\frac{\partial g(x_1-\hat{r}_1)}{\partial(x_1-\hat{r}_1)}\left(\frac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\lambda} - \frac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\mu}\right) + \frac{\partial C(\boldsymbol{t})}{\partial t_\mu} - \frac{\partial C(\boldsymbol{t})}{\partial t_\lambda}\right] = \text{Positive}$$

Where to derive $\dfrac{\partial U(x_1^*(t))}{\partial t_\lambda}$ and $\dfrac{\partial U(x_1^*(t))}{\partial t_\mu}$ we used the envelop theorem. The sign of the expression is positive because $g(x_1 - r)$ is increasing, $\dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\lambda} - \dfrac{\partial \hat{r}_1(\boldsymbol{t})}{\partial t_\mu} > 0$ for Advantaged dictators by Corollary 2.1. Moreover, $C(.)$ is increasing in $|t_k - \bar{t}_k| \forall k \in K$ and, in $\boldsymbol{t}$, $t_\mu > \bar{t}_\mu$ while $t_\lambda < \bar{t}_\lambda$, hence $\dfrac{\partial C(\boldsymbol{t})}{\partial t_\mu} > 0$ while $\dfrac{\partial C(\boldsymbol{t})}{\partial t_\lambda} < 0$.

As such, every $\boldsymbol{t} \in S \ : \ t_\mu > \bar{t}_\mu \wedge t_\lambda < \bar{t}_\lambda$ cannot be optimal for Advantaged dictators.

Also $\boldsymbol{t} \in S \ : \ t_\mu = \bar{t}_\mu \wedge t_\lambda = \bar{t}_\lambda$ cannot be optimum. In fact, $t_\mu = \bar{t}_\mu \wedge t_\lambda = \bar{t}_\lambda$ implies $t_\eta = \bar{t}_\eta$ and, hence, $\boldsymbol{t} = \bar{\boldsymbol{t}}$. From Lemma 5, we know that $\boldsymbol{t}^* \neq \bar{\boldsymbol{t}}$.

Summing up, Expression A.7 is true because we have just excluded the complementary case ($t_\mu^{*A} - \bar{t}_\mu \geq 0 \wedge t_\lambda^{*A} - bart_\lambda \leq 0$) and because Lemma 5 assures the existence of an optimal $\boldsymbol{t}$. As a consequence $t_\mu^{*A} - t_\lambda^{*A} < \bar{t}_\mu - \bar{t}_\lambda$.

The proof for $\bar{t}_\mu - \bar{t}_\lambda < t_\mu^{*D} - t_\lambda^{*D}$ involves showing that

$$(\bar{t_\mu} - t_\mu^{*D} < 0 \wedge \bar{t_\lambda} - t_\lambda^{*D} \geq 0) \vee (\bar{t_\mu} - t_\mu^{*D} \leq 0 \wedge \bar{t_\lambda} - t_\lambda^{*D} > 0) \tag{A.8}$$

As the proof follows the same steps as the proof for the Advantaged case, it is omitted for reasons of space. □

### 3..3 Proof of Corollary 1.1

*Proof.* Let's first look at the Advantaged Dictators. expression A.7 above implies $r(\boldsymbol{t}^*) > r(\bar{\boldsymbol{t}})$. Hence, by Lemma 4 $x_1^*(\boldsymbol{t}^*) > x_1^*(\bar{\boldsymbol{t}})$.

Similarly for Disadvantaged dictators, Expression A.8 above implies $r(\boldsymbol{t}^*) > r(\bar{\boldsymbol{t}})$. Hence, by Lemma 4 $x_1^*(\boldsymbol{t}^*) > x_1^*(\bar{\boldsymbol{t}})$. □

### 3..4 Proof of Proposition 2

*Proof.* Since $\Delta - \text{Attention}^\beta > \Delta - \text{Attention}^\alpha$, $t_\mu^{*\beta} + t_\lambda^{*\beta} = t_\mu^{*\alpha} + t_\lambda^{*\alpha}$, and $\sum_{k\in K} t_k^{*\beta} = \sum_{k\in K} t_k^{*\alpha}$, then:

$$(t_\mu^{*\beta} > t_\mu^{*\alpha} \wedge t_\lambda^{*\beta} \leq t_\lambda^{*\alpha} \wedge t_\eta^{*\beta} = t_\eta^{*\alpha}) \vee (t_\mu^{*\beta} \geq t_\mu^{*\alpha} \wedge t_\lambda^{*\beta} < t_\lambda^{*\alpha} \wedge t_\eta^{*\beta} =_\eta^{*\alpha})$$

The expression above implies $r(\boldsymbol{t}^\beta) < r(\boldsymbol{t}^\alpha)$ for Advantaged dictators and $r(\boldsymbol{t}^{*\beta}) > r(\boldsymbol{t}^{*\alpha})$ for Disadvantaged ones by Corollary 2.1. Hence, Advantaged dictators keep more money for themselves and Disadvantaged ones keep less for themselves by Lemma 4. As a result the Advantaged member of the pair receives more money when $\boldsymbol{t} \in \hat{T}^\alpha$ than if $\boldsymbol{t} \in \hat{T}^\beta$. □

### 3..5 Proof of Proposition 3

*Proof.* We want to prove that under the Assumptions 1, 2, 3, and 4 $E[|x_1^{A-Mer} - x_1^{A-Lib}| > |x_1^{D-Mer} - x_1^{D-Lib}|]$. The proof has three parts. In the first we will rewrite $|x_1^{A-Mer} - x_1^{A-Lib}| > |x_1^{D-Mer} - x_1^{D-Lib}|$ in an expression that depends on the optimal vector of attention when the attention manipulation is not present and on the fairness criteria. The second part of the prove finds an explicit solution for the optimal vector of attention and shows that $|x_1^{A-Mer} - x_1^{A-Lib}| > |x_1^{D-Mer} - x_1^{D-Lib}|$ is positive if and only if $\dfrac{e^A}{e^D} > \sqrt{\dfrac{L}{H}}$. An intermezzo provides the intuition behind this intermediate result. The third part of the proof takes the expected value and shows that it is always positive.

**First part of the proof of Proposition 3** This part of the proof is dedicated to rewrite $|x_1^{A-Mer} - x_1^{A-Lib}| > |x_1^{D-Mer} - x_1^{D-Lib}|$ into an expression that depends on the fairness criteria and the dictators' vector of attention.

*Intermediate step 1.* The attention manipulation shifts the dictator's vector of attention, which in turn shifts $r$, the amount that the dictator considers fair to keep. As a first step, let's check how $x_1$ changes with $r$.

From the proof of Lemma 4 we know that if $x_1 < W$ (which is the relevant case under Assumption 3):

$$\frac{\partial^2 u(x_1)}{\partial^2 x_1}\frac{\partial x_1}{\partial r(\boldsymbol{t})} = \frac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r(\boldsymbol{t}))}\left[\frac{\partial x_1}{\partial r(\boldsymbol{t})} - 1\right]$$

$$\frac{\partial x_1}{\partial r(\boldsymbol{t})} = \frac{-\dfrac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r(\boldsymbol{t}))}}{\dfrac{\partial^2 u(x_1)}{\partial^2 x_1} - \dfrac{\partial^2 g(x_1 - r)}{\partial^2 (x_1 - r(\boldsymbol{t}))}}.$$

From Assumption 1 we have $u(x_1) = x_1$. Hence,

$$\frac{\partial x_1}{\partial r} = 1.$$

As such, the effect of the manipulation on $r$ are translated one to one $x_1$. That is:

$$|x_1^{A-Mer} - x_1^{A-Lib}| = |r^{A-Mer} - r^{A-Lib}|.$$

Where $r^{A-Mer}$ and $r^{A-Lib}$ are the value of $r$ for the Advantaged dictator after restricting the time she can spend on the Merit and Outcome information, respectively. The same is, of course, true for the Disadvantaged dictators.

*Intermediate step 2.* Let's check how the attention manipulation affects $r$. As an example, we compute $r^* - r^{A-Mer}$. By Assumption 3, the manipulation reduces the time the dictator spends on the Merit info by $t_\mu^* - \hat{t}$, and it increases the time she spends contemplating information relevant for the libertarian and egalitarian criteria by $\frac{1}{2}(t_\mu^* - \hat{t})$. This means that, both with and without the manipulation, the dictator spends a total of time equal to $T$ looking at the information. Hence:

$$r(\boldsymbol{t^*}) - r^{A-Mer} = \frac{t_\lambda^{A*}}{T}x_\lambda^A + \frac{t_\mu^{A*}}{T}x_\mu^A + \frac{t_\eta^{A*}}{T}x_\eta - \frac{t_\lambda^{A*} + \frac{1}{2}(t_\mu^{A*} - \hat{t})}{T}x_\lambda^A - \frac{\hat{t}}{T}x_\mu^A - \frac{t_\eta^{A*} + \frac{1}{2}(t_\mu^{A*} - \hat{t})}{T}x_\eta =$$

$$= \frac{1}{T}(\hat{t} - t_\mu^{A*})\left(+\frac{1}{2}x_\lambda^A - x_\mu^A + \frac{1}{2}x_\eta\right).$$

The term in the first parenthesis indicates the decrease in attention to the meritocratic criterion due to the manipulation. Instead, the term in the second parenthesis indicates the change in $r$ per every unit of attention that the manipulation moves away from the meritocratic criterion.

With similar steps we obtain:

$$r(\boldsymbol{t}^*) - r^{A-Lib} = \frac{1}{T}\left(\hat{t} - t_\lambda^{A*}\right)\left(-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta\right)$$

$$r(\boldsymbol{t}^*) - r^{D-Mer} = \frac{1}{T}\left(\hat{t} - t_\mu^{D*}\right)\left(\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)$$

$$r(\boldsymbol{t}^*) - r^{D-Lib} = \frac{1}{T}\left(\hat{t} - t_w^{D*}\right)\left(-x_\lambda^D + \frac{1}{2}x_\mu^D + \frac{1}{2}x_\eta\right).$$

*Intermediate step 3.* We can now study the sign of $r^{A-Mer} - r^{A-Lib}$ and $r^{D-Mer} - r^{D-Lib}$ so that we can rewrite the inequality we want to prove without the absolute value. The total effect of the manipulations on the advantaged is given by:

$$r^{A-Mer} - r^{A-Lib} = [r^{A-Mer} - r(\boldsymbol{t}^*)] - [r^{A-Lib} - r(\boldsymbol{t}^*)] =$$

$$= \frac{1}{T}\left(t_\lambda^{A*} - \hat{t}\right)\left(x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta\right) - \frac{1}{T}\left(t_\mu^{A*} - \hat{t}\right)\left(-\frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta\right) > 0.$$

The last inequality comes from the fact that by Assumption 3 $\bar{t_\lambda} = \bar{t_\mu} = T/3$ and that by the proof of Proposition 1 $t_\mu^{A*} - t_\lambda^{A*} < \bar{t_\lambda} - \bar{t_\mu}$. Hence $\left(t_\lambda^{A*} - \hat{t}\right) > \left(t_\mu^{A*} - \hat{t}\right)$. Moreover, $x_\lambda^A > x_\mu^A$ and so $\left(x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta\right) > \left(-\frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta\right)$.

Using the same steps as above, one can easily prove that:

$$r^{D-Mer} - r^{D-Lib} = [r^{D-Mer} - r(\boldsymbol{t}^*)] - [r^{D-Lib} - r(\boldsymbol{t}^*)] =$$

$$= \frac{1}{T}\left(t_w^{D*} - \hat{t}\right)\left(x_\lambda^D - \frac{1}{2}x_\mu^D - \frac{1}{2}x_\eta\right) - \frac{1}{T}\left(t_\mu^{D*} - \hat{t}\right)\left(-\frac{1}{2}x_\lambda^D + x_\mu^D - \frac{1}{2}x_\eta\right) < 0$$

*Putting together the steps so far.* Using the results of the intermediate steps above we can

rewrite:

$$
\begin{aligned}
&|x_1^{A-Mer} - x_1^{A-Lib}| - |x_1^{D-Mer} - x_1^{D-Lib}| = \\
=&\frac{1}{T}\left(t_\lambda^{A*} - \hat{t}\right)\left(x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta\right) - \frac{1}{T}\left(t_\mu^{A*} - \hat{t}\right)\left(-\frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta\right) + \\
&+ \frac{1}{T}\left(t_w^{D*} - \hat{t}\right)\left(x_\lambda^D - \frac{1}{2}x_\mu^D - \frac{1}{2}x_\eta\right) - \frac{1}{T}\left(t_\mu^{D*} - \hat{t}\right)\left(-\frac{1}{2}x_\lambda^D + x_\mu^D - \frac{1}{2}x_\eta\right) \\
=&\frac{1}{T}(t_\lambda^{A*} - t_\lambda^{D*} + t_\lambda^{D*})\left(x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta\right) + \frac{1}{T}t_\lambda^{D*}\left(x_\lambda^D - \frac{1}{2}x_\mu^D - \frac{1}{2}x_\eta\right) + \\
&- \frac{1}{T}\hat{t}\left(x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta + x_\lambda^D - \frac{1}{2}x_\mu^D - \frac{1}{2}x_\eta\right) + \\
&- \frac{1}{T}\left(t_\mu^{D*} - t_\mu^{A*} + t_\mu^{A*}\right)\left(-\frac{1}{2}x_\lambda^D + x_\mu^D - \frac{1}{2}x_\eta\right) - \frac{1}{T}t_\mu^{A*}\left(-\frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta\right) + \\
&+ \hat{t}\left(-\frac{1}{2}x_\lambda^D + x_\mu^D - \frac{1}{2}x_\eta - \frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta\right)
\end{aligned}
$$

$x_\lambda^A + x_\lambda^D = W$ and $x_\mu^A + x_\mu^D = W$. As such $-\frac{1}{2}x_\lambda^D + x_\mu^D - \frac{1}{2}x_\eta - \frac{1}{2}x_\lambda^A + x_\mu^A - \frac{1}{2}x_\eta = 0$ and $x_\lambda^A - \frac{1}{2}x_\mu^A - \frac{1}{2}x_\eta + x_\lambda^D - \frac{1}{2}x_\mu^D - \frac{1}{2}x_\eta = 0$. Hence we can rewrite the last expression as:

$$
\frac{1}{T}\left[(t_\lambda^{D*} - t_\lambda^{A*})\left(-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta\right) - \left(t_\mu^{A*} - t_\mu^{D*}\right)\left(+\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)\right] \tag{A.9}
$$

The first product in A.9 is the reduction in the amount the Advantaged dictators keep when they are forced to look at the libertarian criterion for the amount of time that is optimal for the Disadvantaged dictators. Vice versa, the second product indicates the reduction in the amount the Disadvantaged dictators when they are forced to look at the meritocratic criterion for the amount of time that is optimal for the Advantaged dictators. Hence, to prove that the behavioral effect of our attention manipulation is stronger for the Advantaged dictators, we need to show that the first effect is stronger than the second one: A.9 must be larger than zero.

**Second part of the proof of Proposition 3.** To check when Expression A.9 is positive we need first to solve the dictator maximization problem to find an explicit solution for $\boldsymbol{t^*}$.

*Finding the optimal vector of attention* $\boldsymbol{t^*}$. Under Assumption 1 and 2 the dictator's utility

function specified in Equation A.1 is:

$$U(x,\boldsymbol{t}) = x_1 - \frac{\beta}{2}(x_1 - \left(\frac{t_\lambda}{t_\lambda+t_\mu+t_\eta}x_\lambda + \frac{t_\mu}{t_\lambda+t_\mu+t_\eta}x_\mu + \frac{t_\eta}{t_\lambda+t_\mu+t_\eta}x_\eta\right)^2 + \\ - \frac{\gamma}{2}\left[(t_\lambda - \bar{t_\lambda})^2 + (t_\mu - \bar{t_\mu})^2 + (t_\eta - \bar{t_\eta})^2\right]$$

As before, we solve the problem sequentially. First the dictator finds the optimal amount to keep $x_1^*$ as a function of $\boldsymbol{t}$. Then she chooses the optimal $\boldsymbol{t^*}$. So the first step is finding:

$$x_1^* = \underset{x_1\in[0;X]}{\arg\max}\, x_1 - \frac{\beta}{2}\left(x_1 - \left(\frac{t_\lambda}{t_\lambda+t_\mu+t_\eta}x_\lambda + \frac{t_\mu}{t_\lambda+t_\mu+t_\eta}x_\mu + \frac{t_\eta}{t_\lambda+t_\mu+t_\eta}x_\eta\right)\right)^2$$

The first order conditions is given by

$$1 = \frac{\beta(x_\eta t_\eta + x_\lambda t_\lambda + x_\mu t_\mu - t_\eta x - t_\lambda x - t_\mu x)}{t_\eta + t_\lambda + t_\mu}$$

From which, noting that the maximand is concave and that the problem has an interior solution according to Assumption 3:

$$x_1^* = \frac{t_\eta + t_\lambda + t_\mu + \beta x_\eta t_\eta + \beta x_\lambda t_\lambda + \beta x_\mu t_\mu}{\beta(t_\eta + t_\lambda + t_\mu)} \tag{A.10}$$

We can now feed the expression for $x_1^*$ in the utility function and find $\boldsymbol{t^*}$.

$$\boldsymbol{t^*} = \underset{\boldsymbol{t^*}\in[0;T]x[0;T]x[0;T]}{\arg\max}\; x_1^* - \frac{\beta}{2}(x_1^* - \left(\frac{t_\lambda}{t_\lambda+t_\mu+t_\eta}x_\lambda + \frac{t_\mu}{t_\lambda+t_\mu+t_\eta}x_\mu + \frac{t_\eta}{t_\lambda+t_\mu+t_\eta}x_\eta\right)^2 + \\ - \frac{\gamma}{2}\left[(t_\lambda - \bar{t_\lambda})^2 + (t_\mu - \bar{t_\mu})^2 + (t_\eta - \bar{t_\eta})^2\right] \\ s.t \\ t_\lambda + t_\mu + t_\eta = T$$

Which gives:

$$t^*_\lambda = \frac{T}{3} + \frac{2x_\lambda - x_\mu - x_\eta}{3T\gamma}, \tag{A.11}$$

$$t^*_\mu = \frac{T}{3} + \frac{-x_\lambda + 2x_\mu - x_\eta}{3T\gamma}, \tag{A.12}$$

$$t^*_\eta = \frac{T}{3} + \frac{-x_\lambda - x_\mu + 2x_\eta}{3T\gamma}. \tag{A.13}$$

The optimal time the dictator spends on a criterion is equal to the value of the bottom up vector of attention plus the attention distortion due to the top-down attention processes. The distortion for one criterion is increasing in the amount that criterion says it is fair to keep, but it is decreasing in the amount the other criteria say it is fair. Intuitively, the gains from distorting attention increase in the distance between criteria. Moreover, the weight of a criterion is equal to 2 in the solution for the optimal time spent on that criterion and -1 in the solution for the other two criteria. The sum of the weight is zero because the time added to one criterion needs to be taken from the other two; the two negative weights are equal to each other due to the symmetry of the cost of distorting attention. Finally, the size of the attention distortion decreases in the $\gamma$ and in $T$. $\gamma$ is the parameter that indicates the relative importance of the cost of manipulating attention, hence an higher $\gamma$ indicates that the attention manipulation is more expensive. $T$ enters in the solution, because the benefit of shifting one unit of attention is decreasing in the total time the dictator can look at the information.

*Substituting the solution for* $\boldsymbol{t^*}$ *in Expression A.9.* We can now go back to check under which conditions Expression A.9 is positive. To do so it is useful to define $e^A$ and $e^D$ as the effort of the Advantaged and Disadvantaged people, respectively. After substituting A.11, A.12, A.13 in Expression A.9, replacing $x^A_\lambda$, $x^D_\lambda$, $x^A_\mu$ and, $x^D_\mu$ with their definitions, and some algebra, we can rewrite Expression A.9 as:

$$\frac{e^A e^D \left[H(e^A)^2 - L(e^D)^2\right](H - L)}{T\gamma(e^A - e^D)^2} \tag{A.14}$$

Which is positive if and only if:

$$\frac{e^A}{e^D} > \sqrt{\frac{L}{H}} \tag{A.15}$$

Inequality A.15 tells us that Expression A.9 is positive as long as the Disadvantaged person does not put too much more effort than the advantaged one (remember $L < H$). Alternatively, we can interpret $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$ as telling us that the Advantaged are so much more productive per unit of effort that the Disadvantaged are far from closing the gap in production even when they put more effort than the Advantaged.

The last results proves Proposition 3. Yet, to get to the intuition behind the result we need more work.

**Intermezzo: building the intuition behind Inequality A.15** We will now try to understand the intuition behind Inequality A.15. To do so, let's start considering the elements of Expression A.9 one by one.

The first product in Expression A.9 is

$$(t_\lambda^{D*} - t_\lambda^{A*}) \left( -x_\lambda^A + \frac{1}{2} x_\mu^A + \frac{1}{2} x_\eta \right)$$

Using A.11, A.12, A.13, the definitions of the fairness criteria, and some algebra, we derive that both terms in the products are positive if

$$H(e^A)^2 - L(e^D)^2 + 3(H - L)e^A e^D < 0. \tag{A.16}$$

As such the product is never negative. To see why the two terms switch sign together consider that

$$sign[-x_\lambda^A + \frac{1}{2} x_\mu^A + \frac{1}{2} x_\eta] = \tag{A.17}$$

$$= - \, sign[2x_\lambda^A - x_\mu^A - x_\eta] = \tag{A.18}$$

$$= sign[-2(X - x_\lambda^D) + (X - x_\mu^D) + (X - x_\eta)] = \tag{A.19}$$

$$= sign[2x_\lambda^D - x_\mu^D - x_\eta]. \tag{A.20}$$

In the second line, the expression in brackets is the numerator of the second term of Equation

A.11 for the case in which the Dictator is *Advantaged*. When this expression is positive, the Advantaged dictators distort their attention towards the libertarian criterion. $\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta < 0$ implies that $2x_\lambda^A - x_\mu^A - x_\eta > 0$ and that these dictators distort their attention towards the libertarian criterion. Instead, in the last line the term in brackets is the numerator of the second term of Equation A.11 for the case in which the Dictator is *Disadvantaged*. Following the same logic as above, we see that $\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta < 0$ implies $2x_\lambda^D - x_\mu^D - x_\eta < 0$ and that the Disadvantaged dictators distort their attention away from the libertarian criterion. As such, the Advantaged and the Disadvantaged dictators always distort their attention towards the libertarian criterion in opposite directions. Moreover, since the bottom-up level of attention to the libertarian criterion does not depend on the dictator status, the sign of $\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta$ also determines the sign of $t_\mu^{A*} - t_\mu^{D*}$.

We can now look at the second product in A.9:

$$\left(t_\mu^{A*} - t_\mu^{D*}\right)\left(+\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right).$$

We the usual substitutions and some algebra, we derive that both terms in this product are positive if and only if

$$H(e^A)^2 - L(e^D)^2 - 3(H-L)e^A e^D > 0. \tag{A.21}$$

Hence, this product is never negative. The intuition behind this result is similar as the one for Inequality A.16. The sign of $\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta$ determines in which direction the dictators distort their attention towards the meritocratic criterion. The direction of the distortion is always opposite for the two types of dictators and hence the sign of $\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta$ determines also the sign of $t_\mu^{A*} - t_\mu^{D*}$.

We established that both products in Expression A.9 are never negative, hence a sufficient condition for A.9 to be positive is that

$$\left(-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta\right) < \left(+\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right) \vee \left(-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta\right) < 0 \tag{A.22}$$

and that

$$t_\lambda^{D*} + t_\mu^{D*} < t_\mu^{A*} + t_\lambda^{A*} \vee t_\lambda^{D*} - t_\lambda^{A*} < 0. \tag{A.23}$$

The second part of Condition A.22 requires that the amount that the Advantage dictators keep

goes down when they are forced to reduce attention to the libertarian criterion by one unit of time. Instead, the first part of the condition requires that this drop in the amount the Advantaged dictators keep is larger than the drop in the amount the Disadvantaged dictators keep when they have to divert attention away from the meritocratic criterion. In other words, restricting by one unit the Advantaged dictators attention to the libertarian criterion should have a larger behavioral effect than restricting by one unit the Disadvantaged dictators attention to the meritocratic criterion.

Instead, the first part of Condition A.23 requires that the Advantaged dictators spend more time looking at information than Disadvantaged ones. This condition is intuitive: for the behavioral effect to be larger for the Advantaged dictators, the attention manipulation should constraint their attention to a larger extent. Finally, the second part of the condition requires that the Advantaged dictators spend more time on the libertarian than on the meritocratic criterion.

*Checking Condition A.22.* Let's start considering the first part of A.22. With the usual substitution and some algebra, we can rewrite

$$\left(-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta\right) < \left(+\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)$$

$$\frac{H(e^A)^2 - L(e^D)^2}{2(e^A + e^D)} > 0$$

The last condition is satisfied when $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$. To see the intuition behind this result, remember that $-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta$ indicates the change in the amount the Advantaged dictator keeps if we restrict by one unit the time she spends thinking about the libertarian criterion. Moreover, let's fix $e^A$ and see how the amounts prescribed by the fairness criteria change when we increase $e^D$. Let's start for the Advantaged dictators. $x_\lambda^A$ remains constant, as it does not depends on $e^D$. $x_\mu^A$ decreases, as the Advantaged can claim a smaller fraction of the money she produced thanks to her random advantaged: $\frac{\partial x_\mu^A}{\partial e^D} = -\frac{(e^A)^2(H-L)}{(e^A+e^D)^2}$. Finally, $x_\eta$ increases because the total pie is going up: $\frac{\partial x_\eta}{\partial e^D} = -\frac{L}{2}$. The speed of decrease of $x_\mu^A$ approximates to zero for high levels of $e^D$, while the $x_\eta$ always increases at a constant speed. Hence, for high enough values of $e^D$, the decrease in $x_\eta$ trumps the increase in $x_\mu^A$. As a consequence the effect of distorting the Advantaged attention away from libertarian criterion moves towards zero and could even become

positive. A similar logic shows that for the Disadvantaged dictators the effect of diverting their attention away from meritocratic criterion are higher for high levels of $e^D$. As the effects of increasing $e^D$ go in opposite directions for the two types of dictators, there must be a single value of $\frac{e^A}{e^D}$ above witch the condition is satisfied.

We now turn our attention to the second part of Condition A.22. It is easy to prove that $-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta < 0$ when $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$. From the derivations above, we know that $-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta$ is positive when Inequality A.16 is satisfied, which is not when $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$. In fact, under this condition the sum of the first two terms of Inequality A.16 is never negative, while the last term of the sum is always positive ($H > L$ by assumption).

*Checking Condition A.23.* Let's now move to the first part of Condition A.23. With the usual substitutions and algebra, we can rewrite:

$$t_\lambda^{D*} + t_\mu^{D*} < t_\mu^{A*} + t_\lambda^{A*}$$
$$-\frac{2}{3}\frac{H(e^A)^2 - L(e^D)^2}{3T\gamma(e^A + e^D)} < 0.$$

Hence, the inequality is satisfied if and only if $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$. To see why this is the condition, we need to go back to the first part of Condition A.22 and consider that

$$\begin{aligned}
&sign[-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta - \left(\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)] = \\
&sign[-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta - \left(\frac{1}{2}(X - x_\lambda^A) - (X - x_\mu^A) + \frac{1}{2}(X - x_\eta)\right)] = \\
&sign[-x_\lambda^A - x_\mu^A + 2x_\eta].
\end{aligned}$$

From Equation A.13 we know that the term in brackets in the last line indicates the sign of the attention distortion for the Advantaged dictators for the egalitarian criterion. A negative sign indicates that these dictators divert their attention towards the egalitarian criterion. When $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$, the first line is negative, and the term in brackets in the last line must be negative as well. Hence, the Advantaged dictators spend less time than $\bar{t} = \frac{T}{3}$ thinking about the egalitarian

criterion.

At the same time, we can also rewrite the same condition as:

$$sign[-x_\lambda^A + \frac{1}{2}x_\mu^A + \frac{1}{2}x_\eta - \left(\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)] =$$
$$sign[-(X - x_\lambda^D) + \frac{1}{2}(X - x_\mu^D) + \frac{1}{2}(X - x_\eta) - \left(\frac{1}{2}x_\lambda^D - x_\mu^D + \frac{1}{2}x_\eta\right)] =$$
$$sign[x_\lambda^D + x_\mu^D - 2x_\eta].$$

From Equation A.13 we know that the term in brackets in the last line indicates the sign of the attention distortion for the Disadvantaged dictators for the egalitarian criterion. A positive sign indicates that these dictators divert their attention towards the egalitarian criterion. When $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$, the first line is negative, and the term in brackets in the last line must be positive. Hence, the Disadvantaged dictators spend more time than $\bar{t} = \frac{T}{3}$ thinking about the egalitarian criterion.

We can now see the intuition behind the result that Condition A.23 is satisfied if and only if $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$. For high value of $\frac{e^A}{e^D}$ the Advantaged Dictators distort their attention away from the egalitarian criterion, while the Disadvantaged dictators distort their attention towards this criterion. All the attention that is not directed to the egalitarian criterion must be allocated to the meritocratic and libertarian ones. Hence the Advantaged dictators spend more time on those criteria than the Disadvantaged. This relationship flips for low enough values of $\frac{e^A}{e^D}$.

*Putting together the different parts of the intuition.* We are now ready to explain the intuition behind Inequality A.15. When the Disadvantaged put less effort or at least not too much more effort than the Advantaged two things happen. First, the behavioral effect of restricting the Advantaged dictators' attention towards the libertarian criterion by one unit of time is larger than restricting the Disadvantaged dictators' attention towards the meritocratic criterion. Second, the Advantaged dictators find more optimal to spend more time on the libertarian and meritocratic criteria than the Disadvantaged dictators. Hence the attention manipulation shifts the Advantaged dictators' attention more than the Disadvantaged dictators' attention. Both effects contributes to ensure that the combined behavioral effect of the attention manipulations is larger for the Advantaged dictators put more effort and, hence, they are more affected by the

manipulations.

**Third part of the proof of Proposition 3.** So far in this proof, we have shown under which condition the attention manipulations change the behavior of the Advantaged dictators more than the behavior of the Disadvantaged ones. We will now show that this differential effect of the attention manipulations holds in expectation as well.

This part of the proof relies on comparing the behavior of the Advantaged and Disadvantaged dictators in two decisions. In the first the Advantaged put more effort than the Disadvantaged. In the second the effort levels are reversed. Formally assume $e, e' : e > e'$. We will compare the way in which the dictators decide when $e^A = e$ and $e^D = e'$ and the way they decide when $e^A = e'$ and $e^D = e$. The key assumption is that the efforts of the Advantaged and Disadvantaged agents are independent and identically distributed. These assumptions assure that the joint probability density function of $e^A = e$ and $e^D = e'$ is the same as joint probability density function of $e^A = e'$ and $e^D = e$. Formally, $f_{e^A,e^D}(e, e') = f_{e^A,e^D}(e', e)$. Hence we can rewrite:

$$
\begin{aligned}
&E\left[|x_1^{A-Mer} - x_1^{A-Lib}| - |x_1^{D-Mer} - x_1^{D-Lib}|\right] = \\
=&E\left[\frac{e^A e^D\left[H(e^A)^2 - L(e^D)^2\right](H-L)}{T\gamma(e^A - e^D)^2}\right] = \\
=&\int_{\underline{e}}^{\bar{e}}\int_{\underline{e}}^{\bar{e}} \frac{e^A e^D\left[H(e^A)^2 - L(e^D)^2\right](H-L)}{T\gamma(e^A - e^D)^2} f_{e^A,e^D}(e^A, e^D) de^A de^D = \\
=&\int_{\underline{e}}^{\bar{e}}\int_{\underline{e}}^{\bar{e}} \frac{ee'\left[H(e)^2 - L(e')^2\right](H-L)}{T\gamma(e-e')^2} + \frac{e'e\left[H(e')^2 - L(e)^2\right](H-L)}{T\gamma(e'-e)^2} f_{e,e'}(e, e') dede' = \\
=&\int_{\underline{e}}^{\bar{e}}\int_{\underline{e}}^{\bar{e}} \frac{ee'(H-L)^2(e^2 + e'^2)]}{T\gamma(e-e')^2} f_{e,e'}(e, e') dede' > 0
\end{aligned}
$$

Where we used Expression A.14 for the first step, and that $H > L$ for the last.

We have now proven that in expectation the attention manipulation changes the behavior of the Advantaged dictators to a larger extent. There is a minority of combinations of $(e^A, e^D)$ for the Disadvantaged dictators put so much more effort than the attention manipulation has larger effects on them. However, in these combinations the difference in behavior is smaller than in the symmetric and equally likely combinations in which the Advantaged dictators put more

effort. □

## 4. Proof of Proposition 4

Here we want to prove that, in expectation, the Advantaged dictators keep more money than the Disadvantaged ones. That is:

$$E(x_1^{*A}) - E(x_1^{*D}) > 0.$$

The first step of the proof is to derive the expression for $x_1^{*A} - x_1^{*D}$. The second is to show that the expected value of this difference is positive.

To get the value of $x_1^{*A} - x_1^{*D}$ we go back to the proof of Proposition 3. There we derived Equation A.10 which gives the amount that is optimal for the dictator to keep. Substituting Equations A.12, A.11, A.13, and the definitions of the fairness criteria in Equation A.10, we find that:

$$x_1^{*A} - x_1^{*D} = \frac{2}{3}\frac{H(e^A)^2 - L(e^D)^2}{e^A + e^D}. \tag{A.24}$$

Equation A.24, indicates that the Advantaged dictators keep more money as long as $\frac{e^A}{e^D} > \sqrt{\frac{L}{H}}$, that is as long as the Disadvantaged dictators did not put much more effort during the production phase.

For the second and last step of this proof, we follow the same logic as in the third step of the proof of Proposition 3. We compare the behavior of the Advantaged and Disadvantaged dictators in two decisions. In the first the Advantaged put more effort than the Disadvantaged. In the second the effort levels are reversed. Formally assume $e, e' : e > e'$. We will compare the way in which the dictators decide when $e^A = e$ and $e^D = e'$ and when $e^A = e'$ and $e^D = e$. We

exploit the fact that $e^A$ and $e^D$ are i.i.d to rewrite:

$$
\begin{aligned}
& E[x_1^{*A}] - E[x_1^{*D}] = \\
=& E\left[\frac{2}{3}\frac{H(e^A)^2 - L(e^D)^2}{e^A + e^D}\right] = \\
=& \int_{\underline{e}}^{\bar{e}} \int_{\underline{e}}^{\bar{e}} \frac{2}{3}\frac{H(e^A)^2 - L(e^D)^2}{e^A + e^D} f_{e^A,e^D}(e^A, e^D) de^A de^D = \\
=& \int_{\underline{e}}^{\bar{e}} \int_{\underline{e}}^{\bar{e}} \frac{2}{3}\frac{He^2 - Le'^2}{e + e'} + \frac{2}{3}\frac{He'^2 - Le^2}{e' + e} f_{e',e}(e, e') dede' = \\
=& \int_{\underline{e}}^{\bar{e}} \int_{\underline{e}}^{\bar{e}} \frac{2}{3}\frac{(H - L)(e^2 + e'^2)}{e + e'} f_{e,e'}(e, e') dede' > 0.
\end{aligned}
$$

The steps above show that in the minority of combinations of $(e^A, e^D)$ in which the Disadvantaged are so much more productive that they keep more money than the Advantaged, the difference in amount kept is smaller that the difference in the symmetric and equally probable case in which the Advantaged dictators are more productive and keeps more. As a consequence the expected value is positive.

# B Additional empirical results

## 1. Introductory Survey Methods and Additional Results

We recruited in total 767 participants from Prolific.co. Because we were interested in how income impacts information-seeking, we separately recruited two groups of about 380 participants at the extremes of income distributions. To do this, we used Prolific filters to recruit participants with personal incomes below 10,000 pounds and household incomes below 50,000 pounds for our low income group (N = 383). For our high income group, we filtered participants to have personal incomes above 70,000 pounds (N = 384). We also restricted our sample to exclude students and to include only participants currently living in the US and UK with an approval rate on Prolific of at least 98%. [25]

In the survey, we first asked participants demographic questions including age, gender, political leaning, current personal income, student status, educational attainment, and car ownership (brand and year). We confirmed participants' student status and personal income by checking whether their answers in our survey matched the prolific filters. If there was a mismatch, participants were informed that they were ineligible for the study and excluded from the survey.

Our main task asked participants to choose between two educational news articles to read. One option focused on success due to merit ("Why high earners work longer hours") and the other focused on the role of luck in success ("Luck looms larger in success than most of us think"). In order to incentivize the choice of articles, participants knew that they could earn a 1 pound bonus by answering comprehension questions about the article correctly on the first try, and they had to answer the questions correctly to proceed. Therefore, they knew that they would have to actually read the article.

After choosing an article and correctly answering comprehension questions, participants were asked about their news consumption behavior, more detailed questions about their source of income, household income and household size, and wealth outside of income, as well as their attitudes toward redistribution. We asked four questions about attitudes toward redistribution that come from the World Values Survey (Haerpfer et al., 2020) and the International Social Survey Programme (ISSP, 2018). These included the share of taxes that high earners should pay, the relative role of luck vs. hard work in success, whether incomes should be more equalized, and governmental vs. individual responsibility. To create an index of redistribution attitudes,

[25] Despite our filtering, a few low income participants reported current household incomes over 50,000 pounds (30/383) and only 8 with incomes over 70,000 pounds. There were even fewer high income participants (8/343) reporting household incomes below 70,000 pounds despite reporting personal incomes of over 70,000 pounds and none who reported household incomes below 50,000 pounds.

we normalized all questions to a range from 1-10 and averaged them.

To check the robustness of our results, we used regressions controlling for demographic variables including age and gender shown in Table A.1. Using a linear probability model to predict article choice, we confirm that high personal income relates to a lower likelihood of choosing the luck article both for binary high compared to low income and a more continuous income measure. A linear regression of our redistribution index again confirms that higher household incomes (binary and continuous) are related to endorsing less redistribution while controlling for demographics.

Table A.1: Survey results.

| | Article Choice | Article Choice | Redistribution Attitude | Redistribution Attitude |
|---|---|---|---|---|
| Constant | $0.38^{***}$ | $0.37^{***}$ | $3.16^{***}$ | $3.18^{***}$ |
| | (0.07) | (0.07) | (0.25) | (0.25) |
| Personal Income (High) | $-0.22^{***}$ | | $1.40^{***}$ | |
| | (0.04) | | (0.14) | |
| Personal Income | | $-0.002^{***}$ | | $0.01^{***}$ |
| | | (0.0003) | | (0.001) |
| Age | $0.005^{**}$ | $0.005^{**}$ | $0.03^{***}$ | $0.03^{***}$ |
| | (0.002) | (0.002) | (0.01) | (0.01) |
| Gender (Woman) | 0.03 | 0.04 | $-0.20$ | $-0.22$ |
| | (0.04) | (0.04) | (0.14) | (0.14) |
| Gender (Other) | 0.12 | 0.13 | $-1.79^{***}$ | $-1.86^{***}$ |
| | (0.12) | (0.12) | (0.48) | (0.48) |
| Gender (Prefer not to answer) | $-0.01$ | $-0.001$ | $-2.70^{*}$ | $-2.76^{*}$ |
| | (0.36) | (0.35) | (1.30) | (1.30) |
| Observations | 767 | 767 | 767 | 767 |

Regressions controlling for demographics including age and gender. Columns (1),(2) show the results of linear probability models with robust standard errors and with dependent variable: article choice where 0 = merit article and 1 = luck article. Columns (3),(4) show the results of linear regressions with dependent variable: attitudes toward redistribution. This is a composite index of four redistribution questions where 1 = more redistribution and 10 = less redistribution. Two measures of personal income are shown, Columns (1) and (3) use binary high or low income, whereas Columns (2) and (4) use a more continuous measure of income in 10,000s of pounds for the high-income participants. Standard errors are shown in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

## 2. Main experiment: recruitment, attrition, and sample characteristics.

**Recruitment procedure.** We recruited our participants on Prolific.co. The study description for the dictators was the following:

*This is a study of the University of Amsterdam on decision making. You will read instructions, answer comprehension questions, complete tasks, and take some decisions. Depending on your and others participants' decisions you might earn a bonus.*
*At the end of the study we will also ask you a few demographic questions. This study consists of two sessions:*

- *Session 1 lasts about 34 minutes*
- *Session 2 lasts about 55 minutes*

*The total reward for completing the two sessions is £9.00:*

- *£2.85 for session 1*
- *£6.15 for session 2*

***You need to complete both sessions for your submission to be approved and paid.***
*You **must** complete Session 2 between Tuesday 24th November at 1.00pm (CET) and Wednesday 25th November at 10.00pm (CET).*
*You **must** use a desktop or a laptop for this study. Submissions from mobile or tablet will be rejected.*
***NOTE:** we do not use deception. Every information we will give you in the instructions is true.*
The recruitment text for the recipient was similar. The main different was that the text did not mention the existence of Day 2.

**Attrition.** Attrition after the treatment assignment was minimal: only 16 dictators (2.6% of the total) started but didn't complete the second day of the experiment, at the beginning of which the treatment was assigned. 6 of these dictators are Advantaged, and 10 are Disadvantaged; Fisher's exact test does not reveal a statistically significant difference in the probability of completing the experiment for these two groups ($p = 0.45$).

**Demographic characteristics.** Table A.2 reports the participant demographics below. These demographics are included as controls in our analyses.

Table A.2: Summary Demographics

| Categories | | % of sample |
|---|---|---|
| Gender | | |
| | Female | 43% |
| | Male | 56% |
| | Other | 1% |
| Age | | |
| | 18 - 29 years | 81% |
| | 30 - 39 years | 13% |
| | 40 - 49 years | 5% |
| | 50 - 64 years | 1% |
| Education | | |
| | Less than high school | 3% |
| | High school | 27% |
| | Some university but no degree | 26% |
| | Bachelor degree | 29% |
| | Master degree | 15% |
| | Doctoral degree | 1% |
| Income | | |
| | Less than 6,000 | 10% |
| | 6,000 - 17,000 | 22% |
| | 17,000 - 35,000 | 31% |
| | 35,000 - 50,000 | 15% |
| | 50,000 - 70,000 | 13% |
| | 70,000 - 87,000 | 5% |
| | 87,000 - 104,000 | 5% |
| Political leaning | | |
| | Left | 16% |
| | Center-left | 33% |
| | Center | 32% |
| | Center-right | 16% |
| | Right | 3% |
| Continent | | |
| | Africa | 2% |
| | Asia | 3% |
| | Europe | 80% |
| | Americas | 16% |

Demographics are summarized for all 600 dictators. Age was reported in years, and reported country was used to determine continent.

## 3. Relationship between Δ Attention and allocations

Figure A.1: The Relationship Between Attention and Allocation.

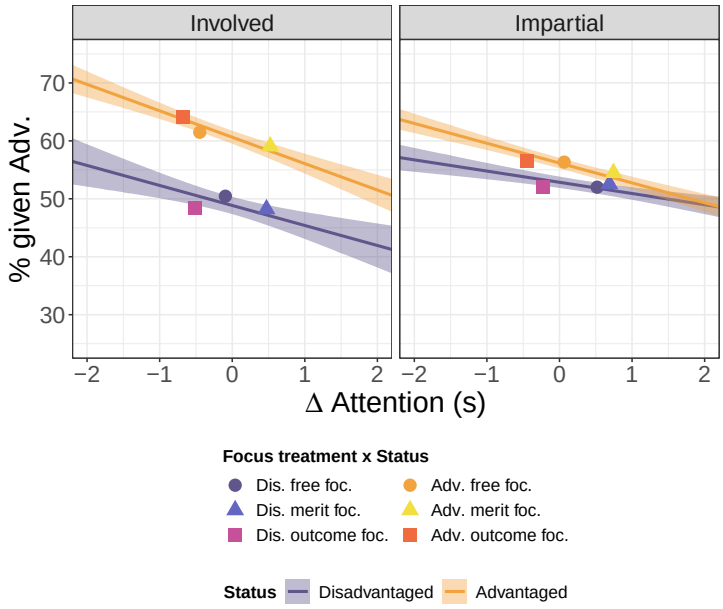



Illustration of the influence of attention on allocations, split by advantaged status. The lines include data from both Free Focus and Constrained Focus treatments and are linearly smoothed with shaded 95% confidence bands. The mean attention and allocations are split by information condition.

Table A.3: The Effect of Attention on Allocation.

| **Panel A: Involved Trials, Constrained Focus only** | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| | All data | Disadvantaged | Advantaged |
| | % to Adv. | % to Adv. | % to Adv. |
| $\Delta$ Attention | -2.63$^{*}$ | -0.089 | -4.09$^{***}$ |
| | (1.21) | (1.91) | (1.17) |
| Observations | 400 | 200 | 200 |
| F-statistic - first stage | 556 | 184 | 406 |
| **Panel B: Impartial Trials, Constrained Focus only** | | | |
| | (1) | (2) | (3) |
| | All data | Disadvantaged | Advantaged |
| | % to Adv. | % to Adv. | % to Adv. |
| $\Delta$ Attention | -0.76 | 0.81 | -1.54$^{+}$ |
| | (0.77) | (1.25) | (0.91) |
| Observations | 400 | 200 | 200 |
| F-statistic - first stage | 245 | 85 | 159 |

The dependent variable is the percentage of the pie allocated to the Advantaged member of the pair. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. The models are instrumental variable regression conducted by 2sls; endogenous regressor: difference in dwell time between merit and outcome information; instrument: Outcome Focus treatment. List of controls: Share of correct answers coming from the advantaged member over the total number of correct answers of the pair, age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 4. The interaction between treatments and the Involved/Impartial condition

Table A.4: Interaction between the between subjects treatments and the Involved/Impartial trials

| | (1) % given Adv. | (2) Δ Attention | (3) % given Adv. |
|---|---|---|---|
| Advantaged | 9.93*** | -0.15* | |
| | (0.98) | (0.076) | |
| impartial | 1.98** | 0.37*** | -0.19 |
| | (0.71) | (0.045) | (0.79) |
| Advantaged * Impartial | -6.52*** | -0.050 | |
| | (0.92) | (0.067) | |
| Outcome Focus | | | 2.73* |
| | | | (1.33) |
| Outcome Focus * Impartial | | | -1.71 |
| | | | (1.24) |
| Observations | 23846 | 23846 | 15881 |

All models are linear regressions. Data: Columns 1 and 2 contain all the observations ; Column 3 excludes the dictators from the Free Focus treatment. Dependent variables: in Columns 1 and 3, the percentage of the pie allocated to the Advantaged member of the pair; in Columns 2: difference in dwell time between merit and outcome information. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls common to all regressions: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories). In addition, Columns 1 and 3 include the share of correct answers coming from the advantaged member over the total number of correct answers of the pair, task type (4 categories).

## 5. Robustness: replicating the main analysis including the demographic and the task type controls

Table A.5: Replicating Table 4 including the controls

| | (1) Inv. trials Δ Attention | (2) Imp. trials Δ Attention | (3) Inv. trals % given Adv. | (4) Imp. trials % given Adv. |
|---|---|---|---|---|
| Free Focus | -0.36 | 0.41 | 27.2*** | 20.8*** |
| | (0.33) | (0.53) | (3.56) | (2.93) |
| Free Focus * Adv. | -0.31$^{+}$ | -0.40 | 9.37*** | 4.24** |
| | (0.17) | (0.25) | (1.58) | (1.28) |
| Merit Focus | 0.25 | 0.64 | 25.8*** | 21.3*** |
| | (0.31) | (0.51) | (3.44) | (2.94) |
| Merit Focus * Adv. | 0.033 | 0.026 | 8.36*** | 1.69 |
| | (0.073) | (0.10) | (1.59) | (1.14) |
| Outcome Focus | -0.75* | -0.22 | 26.5*** | 20.6*** |
| | (0.31) | (0.51) | (3.78) | (3.06) |
| Outcome Focus * Adv. | -0.16* | -0.25* | 12.4*** | 4.47*** |
| | (0.064) | (0.098) | (1.82) | (1.18) |
| Share answers Adv. | | | 65.2*** | 72.3*** |
| | | | (2.12) | (1.79) |
| Observations | 11927 | 11919 | 11927 | 11919 |

All models are linear regressions **without** a constant. Data from the Involved trials in Columns 1 and 3 and from the Impartial trials in Columns 2 and 4. Dependent variable in Columns 1 and 2: difference in dwell time between merit and outcome information. Dependent variable Columns 3 and 4: the percentage of the pie allocated to the Advantaged member of the pair. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls common to all regressions: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories). In addition, Columns 3 and 4 include "Share answers Adv." which is the share of correct answers coming from the advantaged member over the total number of correct answers of the pair, and task type (4 categories).

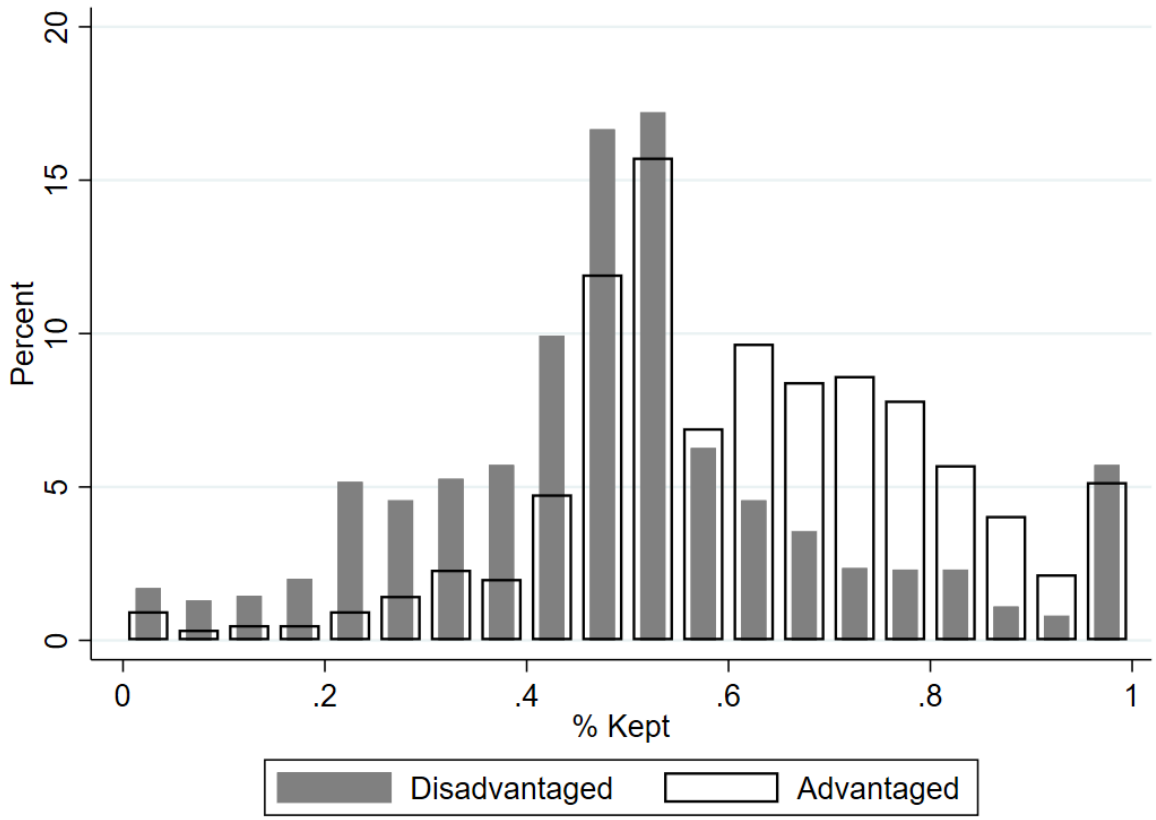
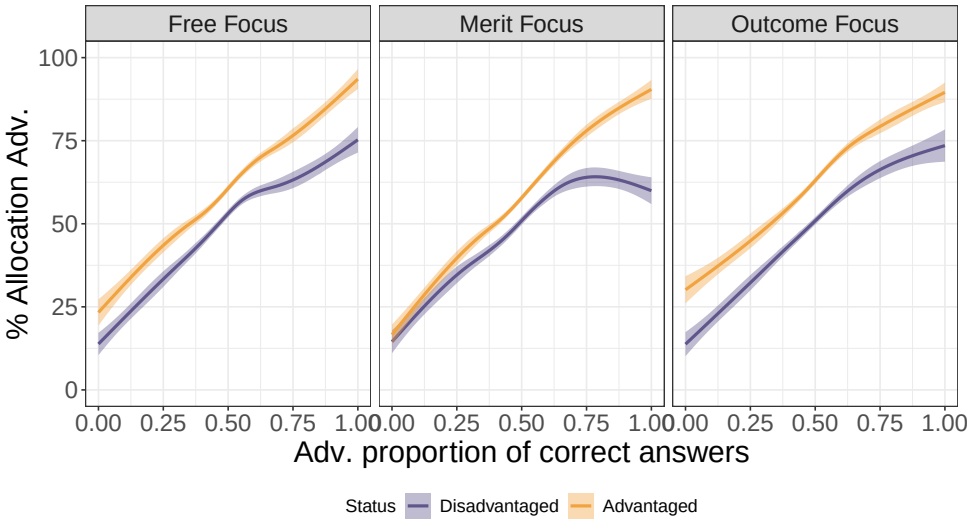



Figure A.2: Histogram of dictators' allocation in the Free Focus treatment divided by Status

## 6. Heterogeneity in allocations

Figure A.3: The responsiveness of allocations to merit




% Allocation Adv. is the percentage of the total pie allocated to the Advantaged member in Involved trials and Adv. proportion of correct answers is the proportion of correct answers (merit) by the Advantaged member of the pair. The relationship between allocations and merit is split by dictator status. The data are displayed with LOESS smoothing and shaded 95% confidence bands.

## 7. Effect of Status on Attention Controlling for Total Dwell Time

Table A.6: Status on Attention - adding controls and total dwell time

| | (1) Δ Attention | (2) Δ Attention | (3) Δ Attention | (4) Δ Attention |
|---|---|---|---|---|
| Advantaged | -0.13$^{+}$ | -0.21$^{*}$ | | |
| | (0.073) | (0.100) | | |
| Free Focus | | | 0.18 | 0.46 |
| | | | (0.33) | (0.52) |
| Free Focus * Adv. | | | -0.29$^{+}$ | -0.39 |
| | | | (0.17) | (0.25) |
| Merit Focus | | | 0.57$^{+}$ | 0.67 |
| | | | (0.31) | (0.51) |
| Merit Focus * Adv. | | | 0.050 | 0.029 |
| | | | (0.079) | (0.10) |
| Outcome Focus | | | -0.40 | -0.19 |
| | | | (0.31) | (0.51) |
| Outcome Focus * Adv. | | | -0.15$^{*}$ | -0.25$^{*}$ |
| | | | (0.060) | (0.098) |
| Observations | 11927 | 11919 | 11927 | 11919 |

This table replicates the results Columns 3 and 4 of Table 3 and of the first two columns of Table 4 including total dwell time as a control.

All models are linear regressions with the dependent variable the difference in dwell time between merit and outcome information. Standard errors clustered by participant in parenthesis. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: total dwell time, age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 8. The interaction effect of Status and round number on Δ Attention

Table A.7: The interaction effect of Status and round number on Δ Attention

| **Panel A: Involved Trials** | | | |
|---|---|---|---|
| | (1) Free Focus Δ Attention | (2) Merit Focus Δ Attention | (3) Outcome Focus Δ Attention |
| Advantaged | 0.078 | 0.092 | -0.11 |
| | (0.18) | (0.085) | (0.073) |
| Round | 0.021* | 0.00040 | 0.017*** |
| | (0.0100) | (0.0048) | (0.0042) |
| Advantaged * Round | -0.037** | -0.0044 | -0.0037 |
| | (0.013) | (0.0062) | (0.0059) |
| Observations | 3985 | 3972 | 3970 |
| **Panel B: Impartial Trials** | | | |
| | (1) Free Focus Δ Attention | (2) Merit Focus Δ Attention | (3) Outcome Focus Δ Attention |
| Advantaged | -0.44+ | 0.13 | -0.14 |
| | (0.26) | (0.11) | (0.11) |
| Round | 0.0033 | 0.010** | 0.011* |
| | (0.0082) | (0.0036) | (0.0042) |
| Advantaged * Round | 0.0059 | -0.0082 | -0.010+ |
| | (0.012) | (0.0055) | (0.0056) |
| Observations | 3980 | 3975 | 3964 |

All models are linear regressions with the dependent variable the difference in dwell time between merit and outcome information. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 9. The interaction effect of Status and attention on Allocations

Table A.8: The Effect of Attention on Allocation - including interaction terms.

| **Constrained Focus only** | | | | |
|---|---|---|---|---|
| | (1)<br>Involved trials<br>% to Adv. | (2)<br>Impartial trials<br>% to Adv. | (3)<br>Involved trials<br>% to Adv. | (4)<br>Impartial trials<br>% to Adv. |
| Outcome Focus | 0.67 | -0.71 | | |
| | (1.93) | (1.12) | | |
| Adv. * Out. Focus | 4.27$^{+}$ | 3.00$^{+}$ | | |
| | (2.33) | (1.59) | | |
| $\Delta$ Attention | | | -0.50 | 0.79 |
| | | | (1.89) | (1.25) |
| $\Delta$ Attention * Adv | | | -3.64$^{+}$ | -2.69$^{+}$ |
| | | | (2.13) | (1.52) |
| Observations | 7942 | 7939 | 400 | 400 |
| F-statistic - first stage | | | 97 | 46 |

The dependent variable is the percentage of the pie allocated to the Advantaged member of the pair. In parentheses, standard errors clustered by participant in Columns 1 and 2, and robust standard errors in Columns 3 and 4. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. Columns 1-2 are linear regressions. Columns 2-3 column are IV regressions. The endogenous regressors are a) the difference in dwell time between merit and outcome information, b) the interaction between this difference in attention and the dummy for the Advantaged status. The instruments are: a) the Focus Treatment, b) the interaction between the Focus Treatment and the Advantaged status dummy. List of controls common to in Columns 1 and 2: Share of correct answers coming from the advantaged member of the pair, task type (4 categories), age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories). List of controls common to in Columns 3 and 4: same controls as the other columns but without task characteristics

## 10. Robustness check: using different dwell time thresholds to disregard the opening of a box

In the main analysis we have considered a box open only if the participant spent at least 200 ms on that box before hovering out because this is the threshold most commonly used (Willemsen and Johnson, 2019; Ghaffari and Fiedler, 2018). However, given that the information displayed to participants is quite simple and there is evidence that information can be processed at shorter timescales, in this appendix we show that the results are robust if we lower the threshold to consider a box open to 100 ms or to 0 ms. The tables below show that changing the threshold neither qualitatively nor quantitatively changes the results.

Table A.9: The Effect of Status on Δ Attention - changing the inclusion criterion for dwell time.

| **Panel A: Average effects of attention** | | | | |
|---|---|---|---|---|
| | Inv. trials<br>Δ Attention 100ms | Imp. trials<br>Δ Attention 100ms | Inv. trals<br>Δ Attention 0ms | Imp. trials<br>Δ Attention 0ms |
| Advantaged | -0.16* | -0.22* | -0.16* | -0.23* |
| | (0.077) | (0.10) | (0.078) | (0.10) |
| Observations | 11927 | 11919 | 11927 | 11919 |
| **Panel B: Interactions between status and attention** | | | | |
| | (1)<br>Inv. trials<br>Δ Attention 100ms | (2)<br>Imp. trials<br>Δ Attention 100ms | (3)<br>Inv. trals<br>Δ Attention 0ms | (4)<br>Imp. trials<br>Δ Attention 0ms |
| Free Focus | -0.35 | 0.42 | -0.36 | 0.42 |
| | (0.33) | (0.54) | (0.34) | (0.54) |
| Free Focus * Adv. | -0.32+ | -0.41 | -0.32+ | -0.42 |
| | (0.17) | (0.26) | (0.18) | (0.26) |
| Merit Focus | 0.28 | 0.66 | 0.28 | 0.66 |
| | (0.31) | (0.51) | (0.31) | (0.51) |
| Merit Focus * Adv. | 0.016 | 0.0088 | 0.012 | 0.0067 |
| | (0.076) | (0.11) | (0.076) | (0.11) |
| Outcome Focus | -0.76* | -0.24 | -0.76* | -0.25 |
| | (0.31) | (0.52) | (0.31) | (0.52) |
| Outcome Focus * Adv. | -0.17** | -0.25* | -0.18** | -0.26* |
| | (0.066) | (0.10) | (0.066) | (0.10) |
| Observations | 11927 | 11919 | 11927 | 11919 |

Panel A replicates Columns 3 and 4 of Table 3. Panel B replicates the first two columns of Table 4. Tables 3 and 4 excluded dwell times below 200ms. Here, Columns 1 and 2 exclude dwell times below 100ms, while Columns 3 and 4 include all dwell times independently on their length.

All models are linear regressions. In Panel A the models include a constant; in panel B the models do not **not** include a constant. Data from the Involved trials in Columns 1 and 3 and from the Impartial trials in Columns 2 and 4. The models use the data from all our dictators. Dependent variable difference in dwell time between merit and outcome information. In Columns 1 and 2 we exclude dwell times shorter than 100ms. In Columns 3 and 4 we don't exclude any dwell time. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

Table A.10: The Effect of Attention on Allocation - changing the inclusion criterion for dwell time.

| **Panel A: Involved Trials, Constrained Focus only** | | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All data | Disadvantaged | Advantaged | All data | Disadvantaged | Advantaged |
| | % to Adv. | % to Adv. | % to Adv. | % to Adv. | % to Adv. | % to Adv. |
| Δ Attention 100ms | -2.52* | -0.085 | -3.94*** | | | |
| | (1.16) | (1.82) | (1.13) | | | |
| Δ Attention 0ms | | | | -2.51* | -0.084 | -3.92*** |
| | | | | (1.16) | (1.82) | (1.12) |
| Observations | 400 | 200 | 200 | 400 | 200 | 200 |
| F-statistic - first stage | 568 | 190 | 415 | 568 | 189 | 415 |
| **Panel B: Impartial Trials, Constrained Focus only** | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | All data | Disadvantaged | Advantaged | All data | Disadvantaged | Advantaged |
| | % to Adv. | % to Adv. | % to Adv. | % to Adv. | % to Adv. | % to Adv. |
| Δ Attention 100ms | -0.74 | 0.77 | -1.50+ | | | |
| | (0.75) | (1.19) | (0.89) | | | |
| Δ Attention 0ms | | | | -0.73 | 0.77 | -1.49+ |
| | | | | (0.74) | (1.18) | (0.89) |
| Observations | 400 | 200 | 200 | 400 | 200 | 200 |
| F-statistic - first stage | 252 | 88 | 164 | 252 | 88 | 164 |

This table replicates Table A.3 changing the inclusion criteria for dwell times. Table A.3 excluded dwell times below 200ms. Here, Columns 1, 2, and 3 exclude dwell times below 100ms, while Columns 4, 5, and 6 include all dwell times independently on their length.

The dependent variable is the percentage of the pie allocated to the Advantaged member of the pair. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. The models are instrumental variable regression conducted by 2sls; Endogenous regressor: difference in dwell time between merit and outcome information; Instrument: Outcome Focus treatment. Data from the Costrained Focus treatments. List of controls: Share of correct answers coming from the advantaged member over the total number of correct answers of the pair, age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 11. The effect of the Outcome Focus treatment on Beliefs

Table A.11: All results in one table Panel A

| | (1) % Disadvantaged that produced more | (2) # recipients with more correct answers |
|---|---|---|
| Outcome Focus | -2.15 | -0.29 |
| | (2.13) | (0.35) |
| Observations | 400 | 400 |

All models are linear regressions. Data:from the Constrained Focus treatments. Robust standard error in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls common to all regressions: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 12. Other measures of Attention

In this section we examine if the instances of looking at information (i.e. the number of times each box is opened) and the last information examined also matter. The number of times each box is opened is similar in concept to dwell time, with the average of each measure per subject highly correlated at 0.88, $p < 0.0001$ for both Involved and Impartial trials. Therefore, it is unsurprising that similar dynamics and a similar relationship between attention and allocations can be found for this measure, illustrated in Figure A.4 and Figure A.5 and Table A.12.

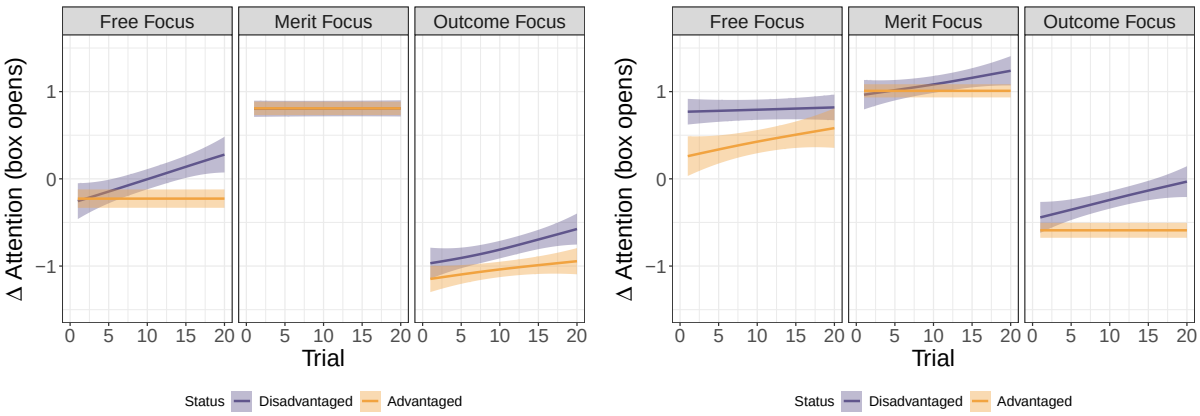
Last fixations are another measure shown to influence choice, as in Ghaffari and Fiedler (2018) who use the last-fixated information as their main attention manipulation. In our data, we find a large difference in this measure between treatments, with the Constrained Focus treatments biasing participants to look at the less-restricted focus information last. We also find differences across Status in the Free and Outcome Focus treatments, with the Disadvantaged participants more often looking at merit information last (Figure A.6 and Table A.13).

Despite the large differences in last fixations, they have a limited effect on allocations in the Constrained Focus treatments. We only find a significant relationship between last-fixated information (merit vs. outcome) and allocations in the Free Focus Involved trials. There, looking last at outcome information increases the allocation to Advantaged participants (Wilcoxon rank-sum test, $p = 0.033$; Merit Focus $p = 0.85$; Outcome Focus $p = 0.48$). This is similar to the findings of Ghaffari and Fiedler (2018) who show that last fixations affect allocation decisions only in self-directed or autonomous attention conditions. They interpret this as an indication of the primacy of preferences rather than salience or other exogenous factors in information search.

Figure A.4: The dynamics of attention (box opens) by treatment

(a) Involved Trials

(b) Impartial Trials




The dynamics of attention (here measured as the number of times merit - outcome information boxes were opened) by round and treatment, in both Involved decisions (left panel) and Impartial decisions (right panel). The data are displayed with LOESS smoothing and shaded 95% confidence bands.

Figure A.5: The relationship between attention (box opens) and allocations

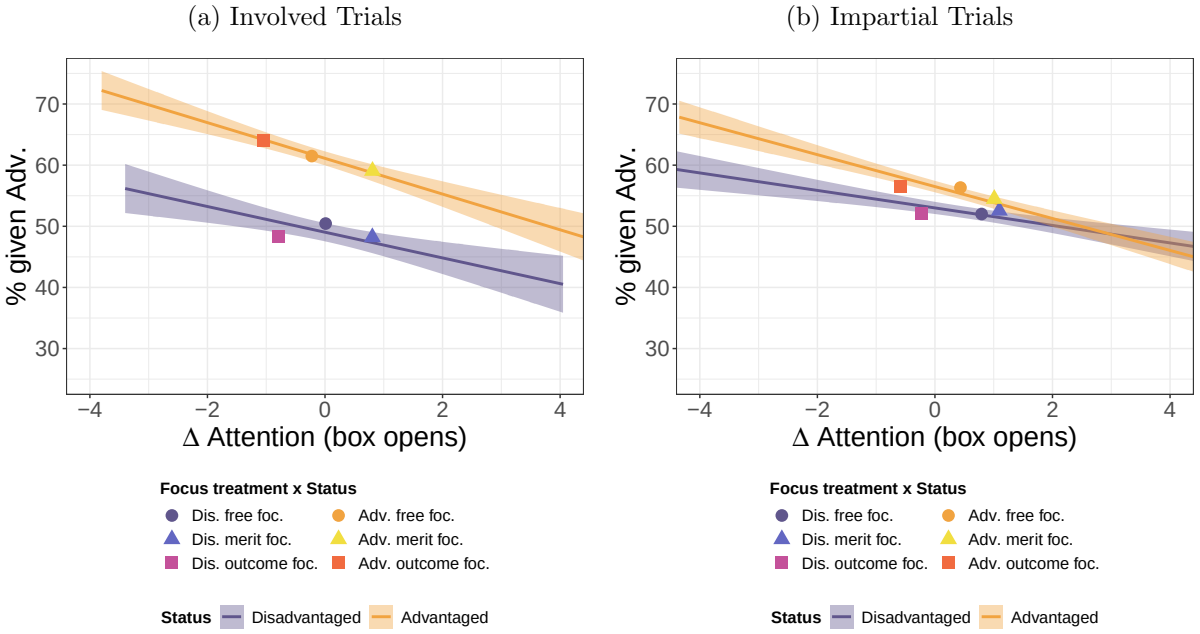
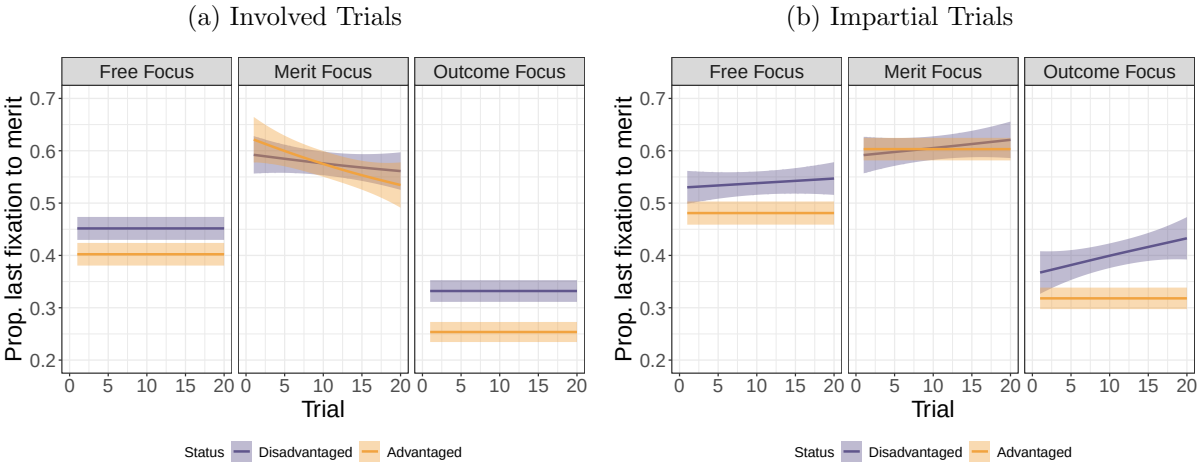



Illustration of the influence of attention (number of times merit - outcome information boxes were opened) on allocations, split by Status. The lines include data from both Free Focus and Constrained Focus treatments and are linearly smoothed with shaded 95% confidence bands. The mean attention and allocations are split by information condition.

Figure A.6: The dynamics of attention (final fixation) by treatment




The dynamics of attention (here measured as the proportion of final fixations on merit information) by round and treatment, in both Involved decisions (left panel) and Impartial decisions (right panel). The data are displayed with LOESS smoothing and shaded 95% confidence bands.

Table A.12: Differences in Attention - Number of Boxes

| | (1) Δ Boxes | (2) Δ Boxes | (3) Δ Boxes | (4) Δ Boxes |
|---|---|---|---|---|
| Advantaged | 0.16 | 0.28* | | |
| | (0.11) | (0.13) | | |
| Free Focus | | | 0.52 | -0.49 |
| | | | (0.40) | (0.53) |
| Free Focus * Adv. | | | 0.17 | 0.28 |
| | | | (0.20) | (0.28) |
| Merit Focus | | | -0.37 | -0.89+ |
| | | | (0.38) | (0.52) |
| Merit Focus * Adv. | | | 0.027 | 0.13 |
| | | | (0.15) | (0.17) |
| Outcome Focus | | | 1.24** | 0.38 |
| | | | (0.39) | (0.51) |
| Outcome Focus * Adv. | | | 0.26* | 0.41* |
| | | | (0.13) | (0.18) |
| Observations | 11927 | 11919 | 11927 | 11919 |

This table replicates the results Columns 3 and 4 of Table 3 and of the first two columns of Table 4 using the difference in number of boxes open containing merit and outcome information as a dependent variable.

All models are linear regressions with the dependent variable difference in number of boxes open containing merit and outcome information as a dependent variable. Columns 3 and 4 do **not** contain a constant. Standard errors clustered by participant in parenthesis. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

Figure A.7: The relationship between attention (final fixation) and allocations

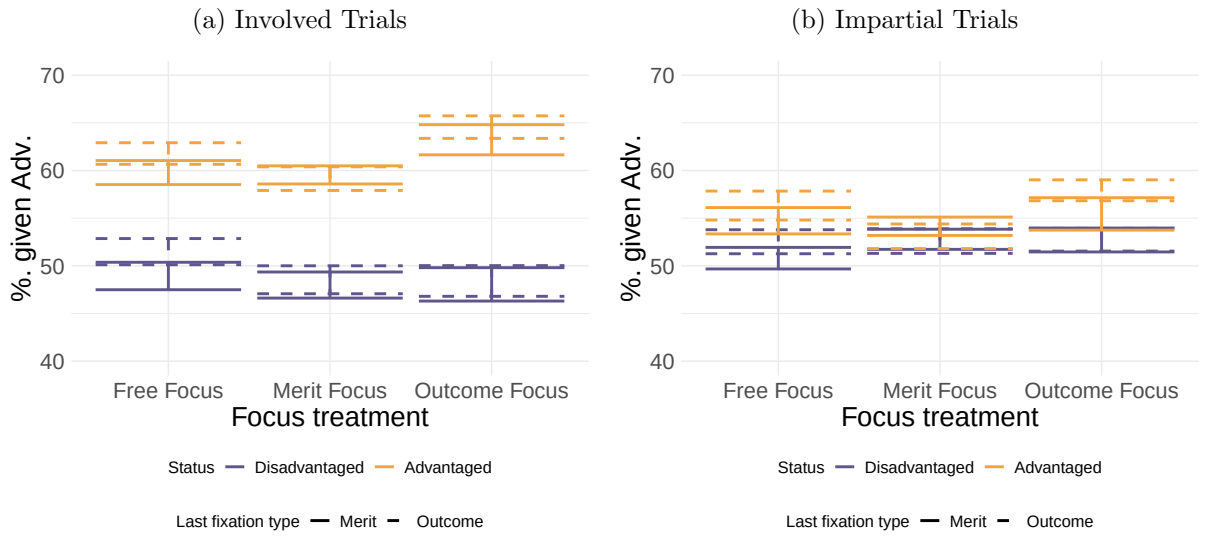



Illustration of the influence of attention (final fixation on merit vs. outcome information) on allocations, split by Status and Focus treatment. 95% confidence intervals are plotted.

Table A.13: Differences in Attention - Last Fixation

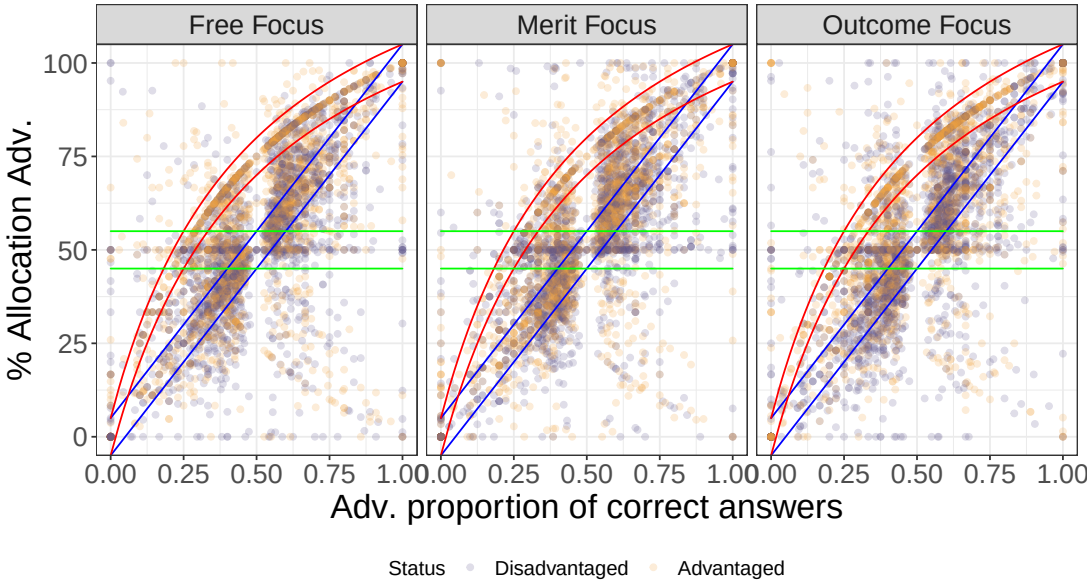
| | (1) Last fixation | (2) Last fixation | (3) Last fixation | (4) Last fixation |
|---|---|---|---|---|
| Advantaged | $0.039^{+}$ | $0.047^{+}$ | | |
| | (0.021) | (0.024) | | |
| Free Focus | | | $0.54^{***}$ | $0.54^{***}$ |
| | | | (0.067) | (0.092) |
| Free Focus * Adv. | | | 0.030 | 0.038 |
| | | | (0.032) | (0.041) |
| Merit Focus | | | $0.41^{***}$ | $0.45^{***}$ |
| | | | (0.068) | (0.090) |
| Merit Focus * Adv. | | | 0.0049 | 0.014 |
| | | | (0.035) | (0.037) |
| Outcome Focus | | | $0.65^{***}$ | $0.65^{***}$ |
| | | | (0.068) | (0.092) |
| Outcome Focus * Adv. | | | $0.081^{*}$ | $0.089^{*}$ |
| | | | (0.033) | (0.040) |
| Observations | 11789 | 11705 | 11789 | 11705 |

This table replicates the results Columns 3 and 4 of Table 3 and of the first two columns of Table 4 the last fixation as a dependent variable.

All models are linear regressions with the dependent variable equal to 1 if the last fixation is on Outcome information. Columns 3 and 4 do **not** contain a constant. Standard errors clustered by participant in parenthesis. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 13. The effect of status and attention of dictators' fairness criteria

Figure A.8: Impartial allocations and fairness criteria




The responsiveness of Impartial allocations to merit and outcome criteria split by Status. Every allocation is plotted as a point, orange for Advantaged and purple for Disadvantaged allocations. The lines represent the boundaries of each criterion (defined as 5% above and below the point-prediction). The red lines represent libertarian criterion, the blue lines represent the meritocratic criterion, and the green lines represent the egalitarian criterion . The gap in allocation data points at exactly 0.50 on the x-axis is due to the fact that we designed trials such that one member always had more correct answers than the other.

Table A.14: Probability of Making a Choice Consistent with a Fairness Criterion

| **Panel A: Egalitarian criterion** | | | |
|---|---|---|---|
| | (1)<br>All Data<br>Egalitarian | (2)<br>Constrained Focus<br>Egalitarian | (3)<br>Constrained Focus<br>Egalitarian |
| Advantaged | -0.063*** | | -0.098** |
| | (0.019) | | (0.032) |
| Outcome Focus | | 0.013 | -0.030 |
| | | (0.024) | (0.035) |
| Adv. * Out. Focus | | | 0.088$^{+}$ |
| | | | (0.047) |
| Observations | 11919 | 7939 | 7939 |
| **Panel B: Meritocratic criterion** | | | |
| | All Data<br>Meritocratic | Constrained Focus<br>Meritocratic | Constrained Focus<br>Meritocratic |
| Advantaged | -0.037$^{+}$ | | 0.037 |
| | (0.021) | | (0.036) |
| Outcome Focus | | -0.027 | 0.043 |
| | | (0.025) | (0.035) |
| Adv. * Out. Focus | | | -0.14** |
| | | | (0.049) |
| Observations | 11919 | 7939 | 7939 |
| **Panel C: Libertarian criterion** | | | |
| | All Data<br>Libertarian | Constrained Focus<br>Libertarian | Constrained Focus<br>Libertarian |
| Advantaged | 0.10*** | | 0.038 |
| | (0.020) | | (0.031) |
| Outcome Focus | | 0.039 | -0.022 |
| | | (0.024) | (0.027) |
| Adv. * Out. Focus | | | 0.12* |
| | | | (0.047) |
| Observations | 11919 | 7939 | 7939 |

Linear probability models for consistency with a fairness norm in the Impartial trials. The dependent variable is a dummy equal 1 if the allocation is consistent with the fairness criterion and zero otherwise. An allocation is considered consistent with a fairness criterion if the distance between the allocation and criterion prescription is less than 5% of the total surplus size. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. Data from the Impartial trials only. List of controls: task type (4 categories), age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

Table A.15: The effect of attention on Fairness Norms - Constrained Focus treatments, separate regressions for Advantaged and Disadvantaged

| **Panel A: Egalitarian criterion** | | | |
|---|---|---|---|
| | (1)<br>Constrained Focus<br>All Data<br>Egalitarian | (2)<br>Constrained Focus<br>Disadvantaged<br>Egalitarian | (3)<br>Constrained Focus<br>Advantaged<br>Egalitarian |
| Outcome Focus | 0.013 | -0.042 | 0.059$^{+}$ |
| | (0.024) | (0.035) | (0.031) |
| Observations | 7939 | 3974 | 3965 |
| **Panel B: Meritocratic criterion** | | | |
| | All Data<br>Constrained Focus<br>Meritocratic | Disadvantaged<br>Constrained Focus<br>Meritocratic | Advantaged<br>Constrained Focus<br>Meritocratic |
| Outcome Focus | -0.027 | 0.051 | -0.093$^{**}$ |
| | (0.025) | (0.034) | (0.036) |
| Observations | 7939 | 3974 | 3965 |
| **Panel C: Libertarian criterion** | | | |
| | All Data<br>Constrained Focus<br>Libertarian | Disadvantaged<br>Constrained Focus<br>Libertarian | Advantaged<br>Constrained Focus<br>Libertarian |
| Outcome Focus | 0.039 | -0.021 | 0.088$^{*}$ |
| | (0.024) | (0.027) | (0.038) |
| Observations | 7939 | 3974 | 3965 |

Linear probability models for consistency with a fairness norm in the Impartial trials. The dependent variable is a dummy equal 1 if the allocation is consistent with the fairness criterion and zero otherwise. An allocation is considered consistent with a fairness criterion if the distance between the allocation and criterion prescription is less than 5% of the total surplus size. Standard errors clustered by participant in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. Data from the Impartial trials of the Constrained Focus treatments only. List of controls: task type (4 categories), age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

Table A.16: The effect of attention of Personal norms - Constrained Focus treatments, separate regressions for Advantaged and Disadvantaged

| **Panel A: Egalitarian criterion** | | | |
|---|---|---|---|
| | (1)<br>Constrained Focus<br>All Data<br>Egalitarian | (2)<br>Constrained Focus<br>Disadvantaged<br>Egalitarian | (3)<br>Constrained Focus<br>Advantaged<br>Egalitarian |
| Outcome Focus | 0.064<br>(0.089) | 0.15<br>(0.13) | 0.0017<br>(0.13) |
| Observations | 400 | 200 | 200 |
| **Panel B: Meritocratic criterion** | | | |
| | All Data<br>Constrained Focus<br>Meritocratic | Disadvantaged<br>Constrained Focus<br>Meritocratic | Advantaged<br>Constrained Focus<br>Meritocratic |
| Outcome Focus | -0.11<br>(0.083) | 0.076<br>(0.12) | -0.27*<br>(0.12) |
| Observations | 400 | 200 | 200 |
| **Panel C: Libertarian criterion** | | | |
| | All Data<br>Constrained Focus<br>Libertarian | Disadvantaged<br>Constrained Focus<br>Libertarian | Advantaged<br>Constrained Focus<br>Libertarian |
| Outcome Focus | 0.016<br>(0.100) | -0.16<br>(0.14) | 0.098<br>(0.14) |
| Observations | 400 | 200 | 200 |

All models are ordered logits. Dependent variable: moral acceptability of a fairness criterion (1 very inappropriate, 2 somewhat inappropriate, 3 somewhat appropriate, 4 very appropriate). Robust standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 14. The effect of attention on Social Norms

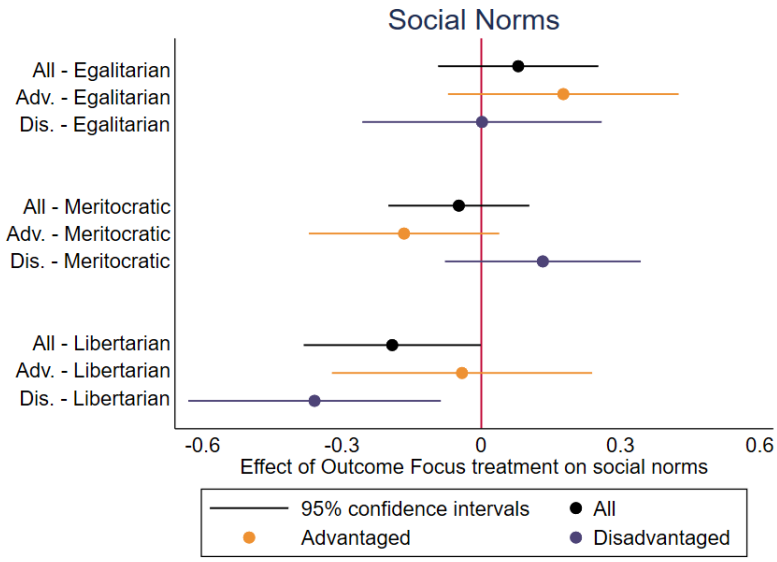



Figure A.9: The graph displays the effect of the Outcome Focus treatment of Social Norms. Data: Merit and Outcome Focus treatments, final questionnaire. Dependent variables: Social Norms, dictators' beliefs about the moral appropriateness of redistributing according to the different fairness criteria. The point estimates indicate the difference between the Outcome Focus and the Merit Focus treatment. The model is an ordered logit. The 95% confidence intervals are computed using robust standard errors clustered at the dictator level.

## 15. The effect of attention on allocation for subjects that did not realize that some boxes where more likely to be restricted than others

Table A.17: Attention on Allocation - including only participants who don't notice any pattern in the attention restrictions

| | (1) Involved trials % given Adv. | (2) Impartial trials % given Adv. | (3) Involved trials % given Adv. | (4) Impartial trials % given Adv. |
|---|---|---|---|---|
| Outcome Focus | 5.06** | 1.81 | 32.7*** | 20.2*** |
| | (1.76) | (1.11) | (7.09) | (4.38) |
| Merit Focus | | | 29.5*** | 20.7*** |
| | | | (6.38) | (4.00) |
| Merit Focus * Adv. | | | 9.04*** | 1.68 |
| | | | (2.17) | (1.29) |
| Outcome Focus * Adv. | | | 12.3*** | 6.17*** |
| | | | (2.32) | (1.71) |
| Observations | 4867 | 4871 | 4867 | 4871 |

This table replicates the results of Columns 5 and 6 of Table 3 and of the last two columns of Table 4 including only the subjects that didn't notice any pattern in the attention restrictions.

All models are linear regressions with the dependent variable the share of the pie given to the Advantaged member of the pair. Data from the Constrained focus treatments. Involved trials in Columns 1 and 3, Impartial trials in Columns 2 and 4. Standard errors clustered by participant in parenthesis. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls: Share of correct answers coming from the advantaged member of the pair, task type (4 categories), age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 16. Within-Subject Variation in Attention Restrictions

Table A.18: Percentage given to the advantaged conditional of the trial and focus type.

| **Involved Trials** | | |
|---|---|---|
| | (1) | (2) |
| | % given to Adv. | % given to Adv. |
| Advantaged | 10.4*** | 10.4*** |
| | (1.29) | (1.29) |
| Outcome Focus | 2.80* | 2.90* |
| | (1.26) | (1.29) |
| OutcomeTrial | | -0.16 |
| | | (0.43) |
| SelfTrial | | -0.19 |
| | | (0.52) |
| OtherTrial | | 0.84$^{+}$ |
| | | (0.51) |
| Observations | 7942 | 7942 |

The dependent variable is the percentage of the pie allocated to the Advantaged member of the pair. Standard errors clustered by participant in brackets. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. Data from the involved trials of the Outcome Focus and Merit Focus treatments. All models are linear regressions. The second columns controls for the round level attention manipulation. Baseline: Merit Trials where we encouraged participants to look at the merit information. In the Outcome Trials participants were encouraged to look at the Outcome information; in the Self Trials they were encouraged to look at the information about their-own monetary contribution and correct answers; in the Other Trials they were encouraged to look at the information about the other player's monetary contribution and correct answers. List of controls: Share of correct answers coming from the advantaged member over the total number of correct answers of the pair, task type (4 categories), age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories).

## 17. Additional information avoidance results

We further quantify information avoidance in Table A.19. Panel A shows the fraction of Involved trials in which subjects open some information, but avoid one type of information (merit or outcome) entirely and Panel B shows avoidance in Impartial trials.[26] While information avoidance levels are overall low and do not drive our main findings (see Tables A.20 and A.21 below for our results excluding trials with avoidance), we can explore which types of information are more likely to be avoided and whether that relates to Status. Disadvantaged dictators are more likely to avoid outcome information (rank-sum test $p = 0.027$), which is in line with them ignoring the luck component in deciding their allocations. However, Disadvantaged dictators are also more likely to avoid merit information (rank-sum test $p = 0.034$) whereas if avoidance were purely self-serving, we would expect higher avoidance of merit by Advantaged dictators. This overall higher avoidance by Disadvantaged dictators may be due to the fact that they are also more likely to endorse and decide in accordance with egalitarian norms, which do not require either type of information, and suggest avoidance is not a mechanism of self-serving bias by Advantaged dictators.

Table A.19: Avoidance of Merit and Outcome Information

| **Panel A: Involved Trials** | | | | | | |
|---|---|---|---|---|---|---|
| | Free Focus | | Merit Focus | | Outcome Focus | |
| | Adv. | Dis. | Adv. | Dis. | Adv. | Dis. |
| Merit avoidance | 4.3% | 4.9% | 1.3% | 6.6% | 2.6% | 2.3% |
| Outcome avoidance | 5.4% | 3.7% | 3.5% | 5.5% | 1.8% | 5.7% |
| **Panel B: Impartial Trials** | | | | | | |
| | Free Focus | | Merit Focus | | Outcome Focus | |
| | Adv. | Dis. | Adv. | Dis. | Adv. | Dis. |
| Merit avoidance | 5.5% | 6.1% | 2.6% | 7.0% | 7.5% | 2.3% |
| Outcome avoidance | 13.8% | 15.6% | 14.0% | 18.7% | 11.0% | 16.9% |

The percentage of trials where a specific type of information is avoided.

In the Impartial trials, where self-serving motivations are removed, there is more avoidance of outcome information, around 15% compared to 5% for merit information, in line with our observation of relatively longer dwell times on merit in Impartial decisions. Furthermore, Disadvantaged participants avoid outcome information more frequently than Advantaged participants

[26] As in all our analyses, this analysis uses filtered data where dwell times of below 200 ms are excluded, which may overestimate avoidance. We exclude trials in which participants did not reveal any information because they are rare and difficult to interpret: it is unclear whether participants were distracted or truly chose not to reveal information on those trials. Such full-avoidance trials account for approximately 1.2% of Involved trials and 1.8% of Impartial trials.

(rank-sum test $p = 0.029$), but there is no difference in avoidance of merit information (rank-sum test $p = 0.31$). This suggests that dictators may not find outcome information as interesting or relevant for Impartial decisions, and in general may show less careful attention as their own payoff is not at stake.

In Tables A.20 and A.21 below we display robustness checks showing that all of our main results hold when excluding trials with ignorance.

Table A.20: Replicating the main results excluding the rounds where dictators don't access all the information.

| | Hypothesis 1<br>% given to Adv. | | Hypothesis 2<br>$\Delta$ Attention | | Hypothesis 3<br>% given to Adv. | |
|---|---|---|---|---|---|---|
| | Involved<br>(1) | Impartial<br>(2) | Involved<br>(3) | Impartial<br>(4) | Involved<br>(5) | Impartial<br>(6) |
| Advantaged | 11.2*** | 3.62*** | -0.15* | -0.16* | | |
| | (1.33) | (0.83) | (0.066) | (0.080) | | |
| Outcome Focus | | | | | 3.13* | 0.62 |
| | | | | | (1.43) | (1.04) |
| Observations | 597 | 572 | 597 | 572 | 398 | 384 |

This Table replicates Table 3 excluding the trials where the dictators don't access all the information and collapsing the data at the individual level.

All models are linear regressions. Data: Columns 1, 3, and 5, Involved trials; Columns 2, 4, and 6, Impartial trials; Columns 5 and 6 exclude the dictators from the Free Focus treatment. Dependent variables: in Columns 1, 2, 5, and 6, the percentage of the pie allocated to the Advantaged member of the pair; in Columns 3 and 4: difference in dwell time between merit and outcome information. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls common to all regressions: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories). In addition, Columns 1, 2, 5, and 6 include the share of correct answers coming from the advantaged member over the total number of correct answers of the pair.

Table A.21: Differences in Share Given to Advantaged

| | Δ Attention | | % given to Adv. | |
|---|---|---|---|---|
| | Involved | Impartial | Involved | Impartial |
| | (1) | (2) | (3) | (4) |
| Free Focus | -0.37 | 0.14 | 37.3*** | 17.9** |
| | (0.24) | (0.37) | (10.4) | (5.58) |
| Free Focus * Adv. | -0.29* | -0.39* | 10.2*** | 4.10** |
| | (0.14) | (0.19) | (1.69) | (1.49) |
| Merit Focus | 0.27 | 0.50 | 35.4*** | 19.2*** |
| | (0.22) | (0.36) | (9.73) | (5.46) |
| Merit Focus * Adv. | -0.027 | 0.051 | 9.98*** | 1.21 |
| | (0.057) | (0.078) | (2.08) | (1.36) |
| Outcome Focus | -0.83*** | -0.50 | 36.5*** | 17.4** |
| | (0.22) | (0.36) | (10.3) | (5.54) |
| Outcome Focus * Adv. | -0.13* | -0.14+ | 13.3*** | 5.58*** |
| | (0.056) | (0.079) | (2.06) | (1.42) |
| Observations | 597 | 572 | 597 | 572 |

This Table replicates Table 3 excluding the trials where the dictators don't access all the information and collapsing the data at the individual level.

All models are linear regressions. The models do **not** include a constant. Data from all dictators, Involved trials in Columns 1 and 3, and Impartial trials in Columns 2 and 4. Dependent variable in Columns 1 and 2: difference in dwell time between merit and outcome information. Dependent variable Columns 3 and 4: the percentage of the pie allocated to the Advantaged member of the pair. Robust standard errors in parentheses. $^{+}$ $p < 0.10$, $^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$. List of controls common to all regressions: age, gender (man, woman, other), political affiliation (5 categories), education (6 categories), income (7 categories), continent (4 categories). Columns 3 and 4 also include the share of correct answers coming from the advantaged member over the total number of correct answers of the pair

# C Preregistration

# CONFIDENTIAL - FOR PEER-REVIEW ONLY

## Tracking fairness (#44417)

**Created:** 07/12/2020 01:54 AM (PT)
**Shared:** 11/03/2020 12:17 PM (PT)



**1) Have any data been collected for this study already?**
No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**
We study the origin of self-serving biases in monetary allocation problems. If people are randomly placed in a (dis)advantaged position, how does this affect their attention to meritocratic information, the ethical criteria for making decisions, and the subsequent allocation choices? Detailed hypotheses are specified in point 5).

**3) Describe the key dependent variable(s) specifying how they will be measured.**
In Part 1 of the experiment, subjects first produce a surplus together with a matched partner on several tasks. We create variation in contribution to the surplus by randomly giving one of the partners a higher piece rate than the other. In Part 2 of the experiment, some subjects are given information on the performance on the tasks as well as the total contribution, and make allocation decisions in the role of dictator. We use Mouselab to track the way subjects explore information about task performance.

Per every decision of the dictator we record:
- the split in the total surplus between dictator and recipient.
- dwelling time (mousetracked) on each of the following information 1) the dictator & recipient contribution to the pie in monetary terms, 2) the number of answers in the task the dictator & recipient got correct.

**4) How many and which conditions will participants be assigned to?**
Subjects are assigned to be “receivers” and “dictators”. Both groups take part in a series of performance tasks to determine the surplus. We are mostly interested in the dictators.

All dictators are assigned to one of two treatments:
Advantaged: receives a high piece rate per correct answer in the task.
Disadvantaged: receives a low piece rate per correct answer in the task.

Each dictator participates (in this order) in an
Involved condition: 20 allocations between themselves and another randomly matched participant
Benevolent condition: 20 allocations between two other participants.

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**
Hypothesis 1 (Behavior): In the involved condition, advantaged dictators give less money to the receivers than disadvantaged dictators.

We test this hypothesis with a non-parametric rank sum test. We will perform regressions to control for subject characteristics with standard errors clustered for each participant.

Hypothesis 2 (Attention): In the involved condition, advantaged dictators spend relatively less time on correct answer information and more time on monetary contribution information than disadvantaged dictators.
Across dictator groups, we investigate total time looking at information as well the proportion of time spent looking at correct answers, using a non-parametric rank sum test. We will also perform regressions with standard errors clustered for each participant.

Hypothesis 3 (Persistence): The effects documented in 1) and 2) persist in the benevolent condition.
The tests are the same as for Hypothesis 1 and 2, but now in the benevolent condition. We will also compare the effects in both conditions using a difference in difference approach.

Hypothesis 4 (Role of attention): Attention patterns drive giving decisions.
For correlational evidence, we use regressions to investigate how sensitive the treatment effect (Hypothesis 1) is to controlling for total and relative looking time. For a causal inference, we use an instrumental variable analysis to exploit variation generated by the (randomly varied) orientation of patterns on the

screen.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**
Following standard Mouselab protocols, we will exclude information that was revealed for less than 200 ms.

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**
We will recruit 200 dictators from the online platform Prolific. These are divided 50-50 between the advantaged and disadvantaged condition. We recruit the corresponding number of recipients.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**
We will conduct a number of secondary analyses:
- We will compare by treatment the fairness criteria people list in the questionnaire as being most socially appropriate.
- We compare by treatment the fairness "types" based on Cappelen et al. (2007), and correlate these types with attentional patterns.
- Correlate attention, behavior and political preferences elicited in the final questionnaire.
In addition, we will explore additional measures of attention, and their explanatory power for giving decisions. We will conduct robustness analysis on the revelation threshold in point 6).

# CONFIDENTIAL - FOR PEER-REVIEW ONLY

## Tracking fairness - attention manipulation (#52512)

**Created:** 11/18/2020 09:42 AM (PT)
**Shared:** 02/16/2021 06:11 AM (PT)



**1) Have any data been collected for this study already?**
No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**
We study the origin of self-serving biases in monetary allocation problems. If people are randomly placed in a (dis)advantaged position, how does this affect their attention to meritocratic information, the ethical criteria for making decisions, and the subsequent allocation choices? In a previous version of the experiment, we showed that advantaged dictators pay less attention to information that reveals pure merit (actual task performance). In this experiment we ask how randomly induced variations in attention affect decision making.

**3) Describe the key dependent variable(s) specifying how they will be measured.**
In Part 1 of the experiment, subjects first produce a surplus together with a matched partner on several tasks. We create variation in contribution to the surplus by randomly giving one of the partners a higher piece rate than the other. In Part 2 of the experiment, some subjects are given information on the performance on the tasks as well as the total contribution, and make allocation decisions in the role of dictator. We manipulate how long different kinds of information are available to people.

Per every decision of the dictator we record:
- the split in the total surplus between dictator and recipient.
- dwelling time (mousetracked) on each of the following information 1) the dictator & recipient contribution to the pie in monetary terms, 2) the number of answers in the task the dictator & recipient got correct.

**4) How many and which conditions will participants be assigned to?**
Subjects are assigned to be “receivers” and “dictators”. Both groups take part in a series of performance tasks to determine the surplus. We are mostly interested in the dictators.

All dictators are assigned to one of two treatments:
Advantaged: receives a high piece rate per correct answer in the task.
Disadvantaged: receives a low piece rate per correct answer in the task.

We cross-randomize these treatments with another dimension:
Merit focus: in a majority of trials, the information about task performance (merit) is available longer.
Output focus: in a majority of trials, information about total contribution to surplus is available longer.

Each dictator participates (in this order) in an
Involved condition: 20 allocations between themselves and another randomly matched participant
Benevolent condition: 20 allocations between two other participants

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**
We test two main hypotheses for both the involved and the benevolent dictators:
1) Dictators in the “Merit Focus” treatment will give more to disadvantaged recipients.

We will test this in a regression with data for all trials and a dummy for all trials with Merit Focus, as well as controls for subject and trial characteristics.

2) Compared to a situation with freely chosen attention, making dictators look longer at “inconvenient” information (i.e. ”Merit focus” for advantaged dictators, “Output focus” for disadvantaged dictators) will reduce the relative bias of advantaged dictators towards the advantaged recipients.

We combine the data from this experiment with a previous experiment in which dictators could freely choose what to look at. We will use regressions to evaluate the “difference in difference”.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**

Following standard Mouselab protocols, we will exclude information that was revealed for less than 200 ms.

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

We will recruit 400 dictators from the Prolific platform. Dictators will be evenly split between the 4 between subject conditions (i.e. 100 in each cell). We recruit a corresponding number of receivers.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

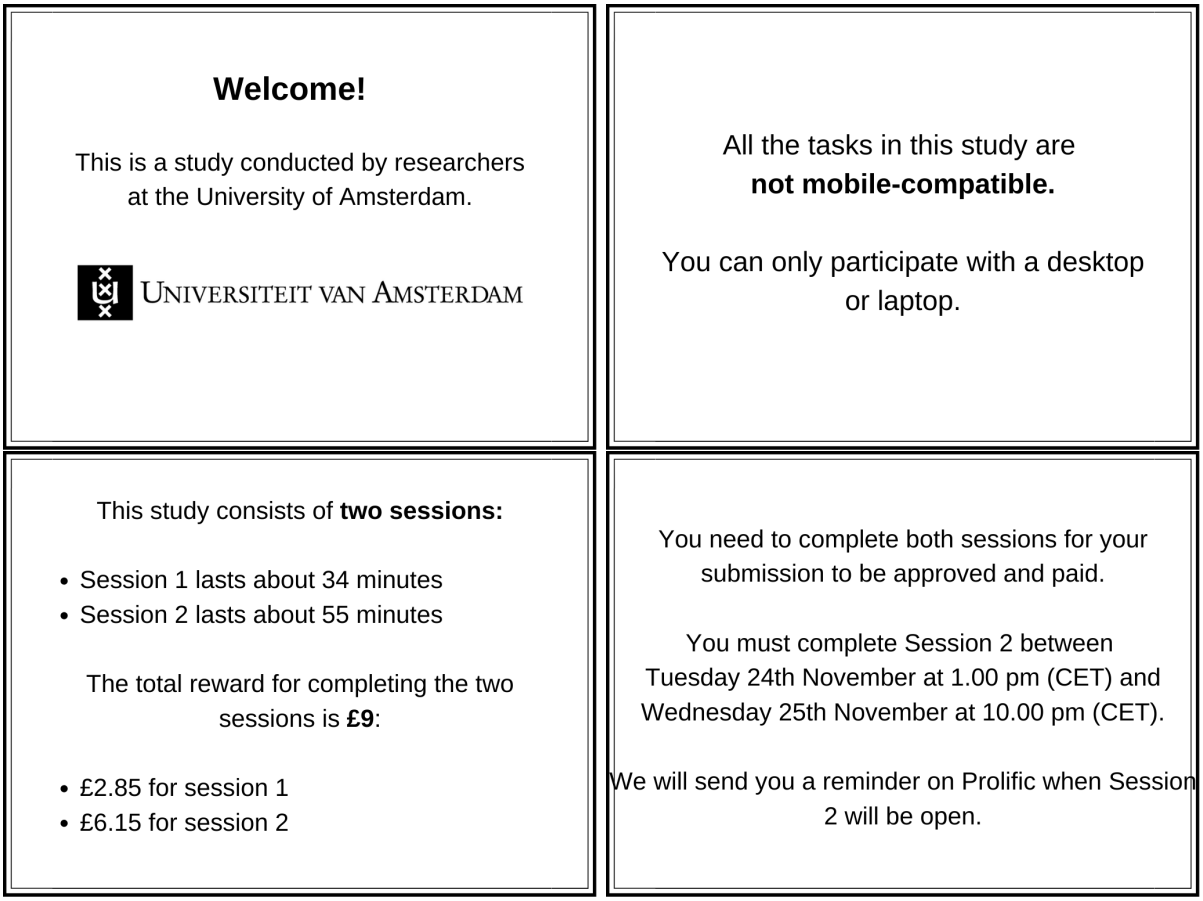
We will investigate whether the impact of merit/output information on giving differs between advantaged and disadvantaged dictators.

We will correlate giving and attention with several additional elicitations in the questionnaire on perceptions of fairness.

# D Instructions

The instructions for the dictators in the experiment are shown below, together with the comprehension questions. The instructions were presented in several decks of slides. Participants could move across slides clicking on two buttons on the sides of the screen. Comprehension questions were presented on a separate page. Participants could move back to the instruction from the page with the comprehension questions.

## 1. Day 1




For the recipients the third slide of this set read: "This study consists of two sessions. However, you will only participate in Session 1 that takes place today. Session 1 lasts about 34 minutes. The reward for completion is £2.85.

On top of the base earnings, in this study you can earn a **bonus.**

The amount of the bonus may depend on:

- Your **performance**
- Your **decisions** or the decisions of another participant
- **Luck**

The Ethics Committee Economics and Business (EBEC) of the University of Amsterdam has approved our study (EC 20200214090248). You can contact our Ethics Committee writing to secbs-abs@uva.nl.

To receive the approval we committed ***not*** to use misleading or untruthful instructions.

UNIVERSITEIT VAN AMSTERDAM

Please answer the questions on the next webpage.

**Comprehension questions:**

1. Your bonus might depend on the decisions taken by another participant. T/F [correct: T]
2. You can complete this study using a mobile device. T/F [F]
3. According to the ethical protocol under which we run this study, all the instructions you read must be truthful and not misleading. T/F [T]
4. You need to complete both sessions of this study for your submission to be approved. T/F [T]

**Instructions for Session 1**

In this session, you will do 8 tasks.

In each task you will have to answer several questions.

**Within each task**, every correct answer gives the same monetary reward.

However, **different tasks** give different monetary rewards per correct answer.

There are two possible reward levels, **high** and **low.**

**We will split participants into two groups:**

**50% in the High Reward Group**

**50% in the Low Reward Group**

**What is the difference?**

Participants in the **High Reward Group** will always receive the **High Reward** per correct answer.

Participants in the **Low Reward Group** will always receive the **Low Reward** per correct answer.

**Luck decides if you are in the High or Low Reward Group.**

The computer determines your group randomly. You have 50% chance to be in either group

If the computer assigns you to the **High Reward Group**, then you will be in the High Reward Group for the **whole study**.

Similarly, if the computer assigns you to the **Low Reward Group**, then you will be in the Low Reward Group for the **whole study**.

You will know if you are in the High or Low Reward Group only at a later stage.

In each task, you are matched with a different participant

After each task, the reward that both you and the other participant earned flows into a common account.

At a later stage, we will explain how the money in this account is paid out to the members of the pair.

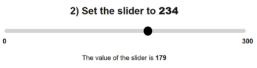
Please answer the questions on the next webpage.

1. In this study you have to complete BLANK tasks. [8]
2. There are 3 groups of participants. T/F [F]
3. Luck determines if you are in the High Reward Group or in the Low Reward Group. T/F [T]
4. In some tasks, you will be in the High Reward Group, in others you will be in the Low Reward Group. T/F [T]

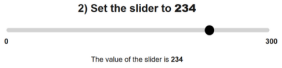
**Tasks 1 and 2**

You need to move a slider to a predetermined value.
You have **2 minutes** to move as many sliders as you can.

**Incorrect slider**




**Correct slider**




You get a reward per each slider you set to the correct value.

**High Reward: £0.30** **Low Reward: £0.10**

**Tasks 3 and 4**

You need to answer a quiz.
You have **4 minutes** to answer as many questions as you can.

You get a reward per each correct answer.

**High Reward: £0.90** **Low Reward: £0.30**

**Tasks 5 and 6**

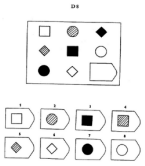
You need to count zeros in a table.
You have **3 minutes** to count the zeros in as many tables as you can.


You get a reward each time you correctly state the number of zeros.

**High Reward: £0.30** **Low Reward: £0.10**

**Tasks 7 and 8**

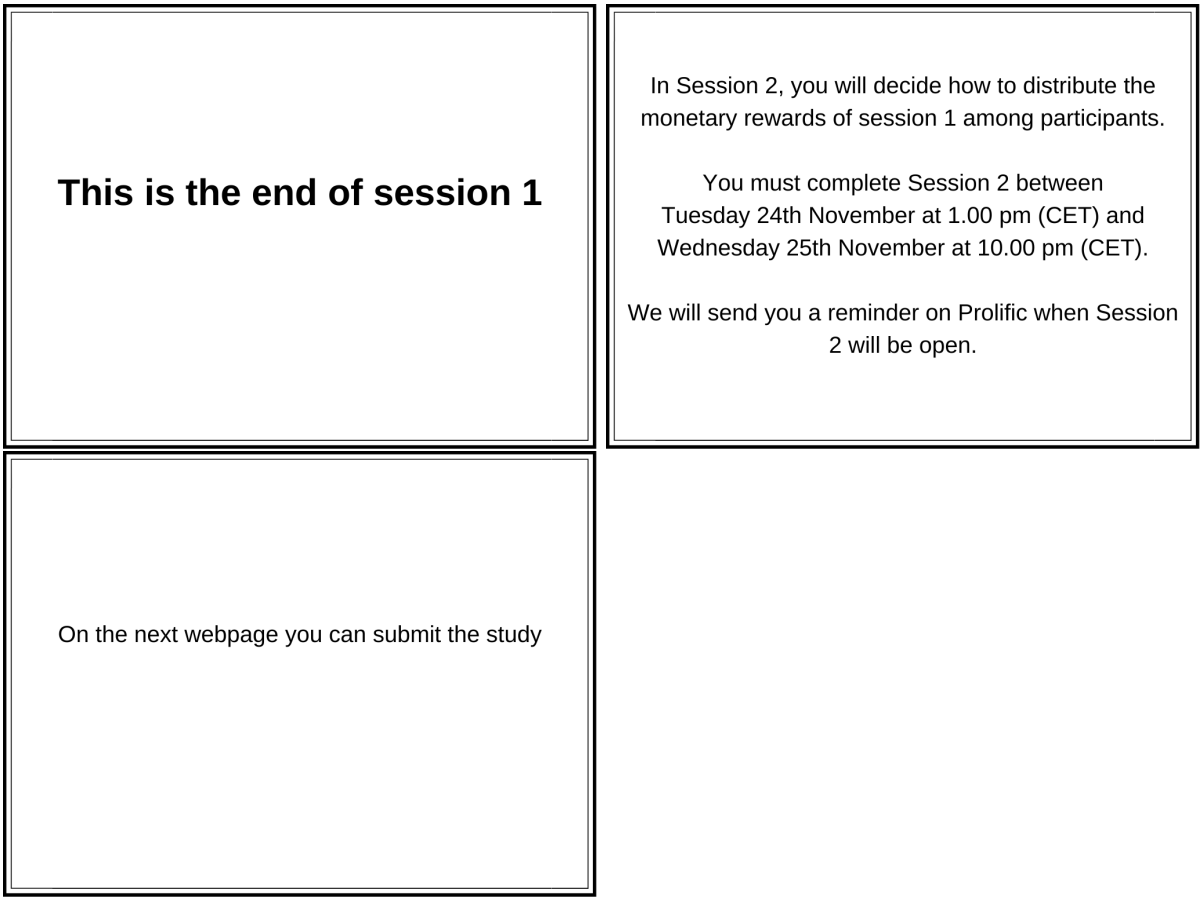
You need to find the missing tile in a figure.
You have **3 minutes** to complete as many figures as you can.



You get a reward each time you select the correct tile.

**High Reward: £0.60** **Low Reward: £0.20**

The slides with the task instruction appeared before the relevant pair of tasks. To continue to the task, the participants had to correctly input the two possible pay-rates for the task.




In this last set of slides for the recipients, the second slide read: "We will approve your submission within two working days. The participants in Session 2 will decide how to distribute the monetary rewards of session 1 among participants. We will let you know your bonus and your pay rate with a Prolific message".

## 2. Day 2

Comprehension questions:

1. I confirm that I am using a laptop or desktop. Y/N [Yes]
2. Your performance on the tasks in Session 1 carries over into Session 2. T/F [True]
3. We commit to providing entirely accurate and truthful information in all aspects of this study. T/F [True]

## Welcome!

This is a study conducted by researchers at the University of Amsterdam.

UNIVERSITEIT VAN AMSTERDAM

All the tasks in this study are **not mobile-compatible.**

You can only participate with a desk- or laptop, **preferably with a mouse** instead of a mousepad.

The Ethics Committee Economics and Business (EBEC) of the University of Amsterdam has approved our study. You can contact our Ethics Committee writing to secbs-abs@uva.nl.

To receive the approval we committed ***not*** to use misleading or untruthful instructions.

UNIVERSITEIT VAN AMSTERDAM

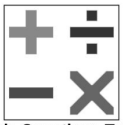
Welcome to Session 2 of this study.

This session consists of three parts that will last for approximately **55** minutes.

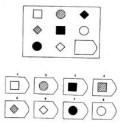
In Session 1, you completed a series of tasks online and you earned a reward for each correct answer.


**Moving Sliders Task**

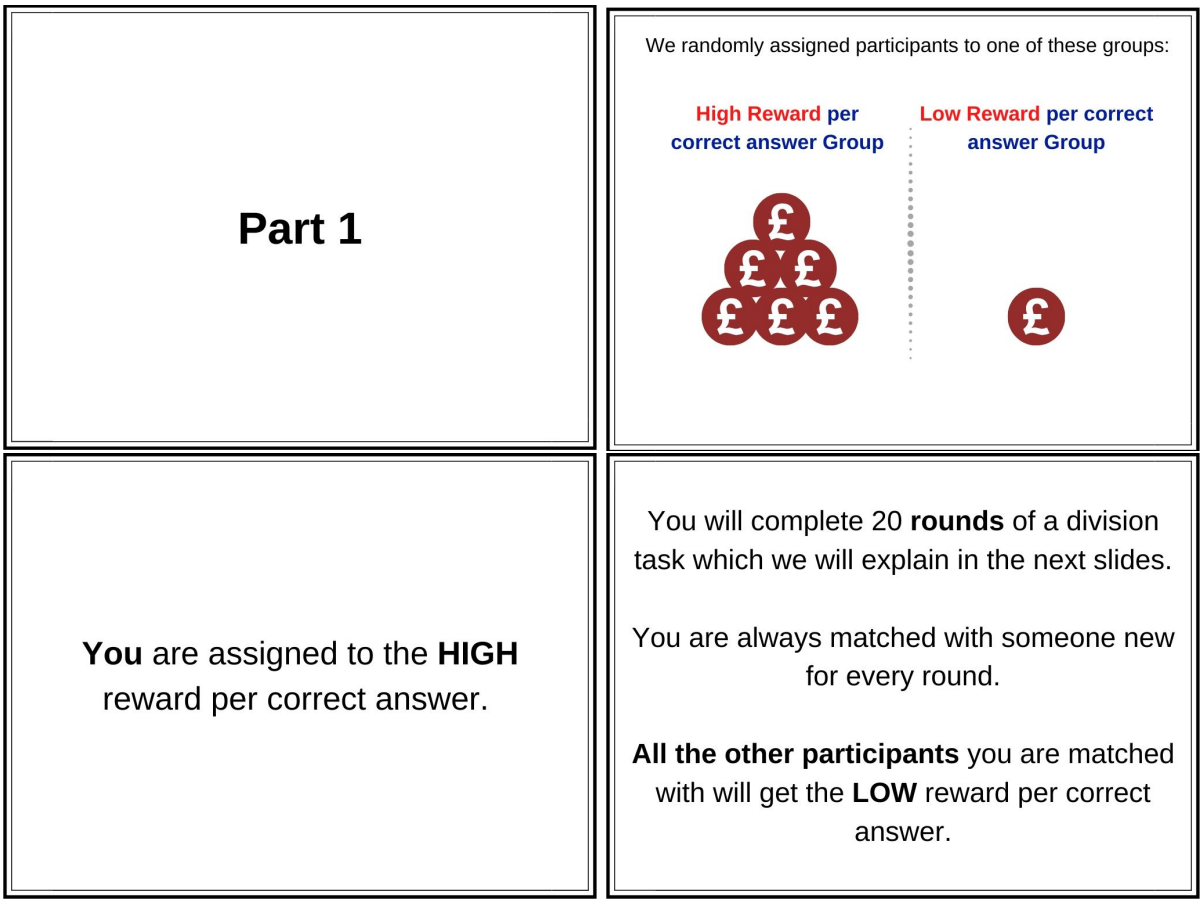
**Counting Zeros Task**



**Quiz Questions Task**



**Missing Tiles Task**

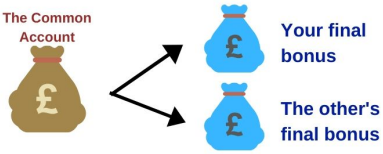



The second row shows the instructions for Advantaged participants, whereas Disadvantaged participants saw HIGH and LOW switched across slides. Disadvantaged participants were instructed that that they were assigned the LOW, and the other participants the HIGH, reward per correct answer.

In every round, the earnings from both participants on one of the tasks are combined into a **common account**.


Your task in Part 1 is to **divide these common accounts** between yourself and the other participants.

Your decision on **one of the rounds** might determine the bonus that you and the other participant receive at the end.




Your final bonus is based on **one randomly selected decision** in Part 1 or Part 2.

If your bonus comes from Part 1, one of the rounds will be randomly selected by the computer.

You and the other participant will earn the bonus determined by your division of the common account in that round.

You will be paid the money you allocated for yourself, and the other participant will receive the money you allocated to them.

**Consider each decision carefully,** as any of them could be selected for payment.

**Comprehension questions:**

1. Which reward condition were you and the other participants you are matched with assigned to? MULTIPLE CHOICE [Advantaged: You: High reward, Others: Low reward; Disadvantaged: You: Low reward, Others: High reward;]

2. What determines the common account on each round? MULTIPLE CHOICE [The combined amount you and the other participant earned on a single task from Session 1]

3. If Part 1 determines the bonus, how will you be paid? MULTIPLE CHOICE [The amount you gave yourself on a randomly selected round from Part 1]

4. If Part 1 determines the bonus, how will the other participant you are matched with be paid? MULTIPLE CHOICE [The amount you gave them on a randomly selected round from Part 1]

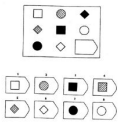
**Before you decide** on the division of the common account, we will show you some **additional information**.

First, you will see **which task** from Session 1 determined the common account for this round.

**Moving Sliders Task**

**Counting Zeros Task**

**Quiz Questions Task**



**Missing TilesTask**

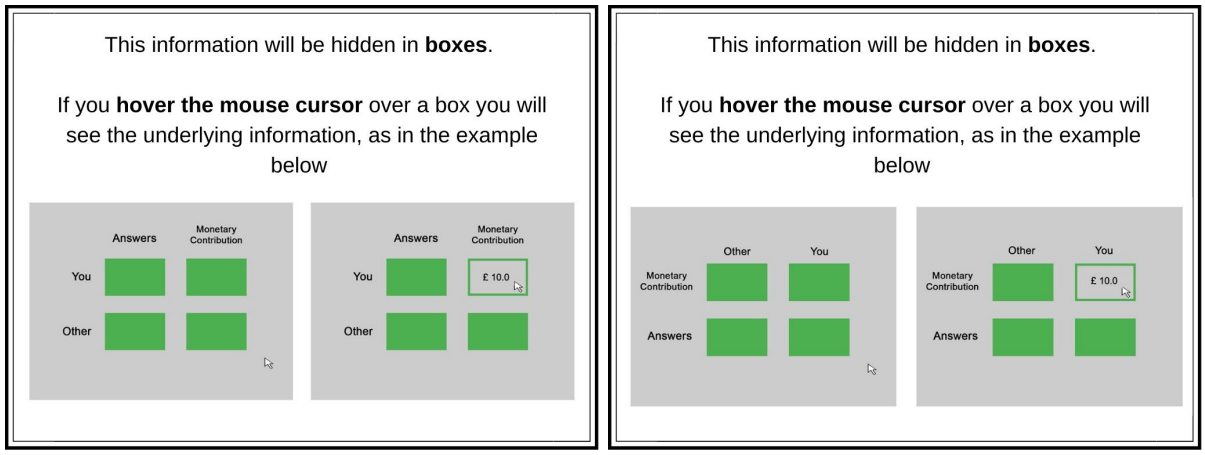
On the same page, you will also see the **total earnings** in the common account that you will divide, as in the example below:

**Task: Counting Zeros**

**Common account: £10**

Next, you have the chance to learn how both participants contributed to the common account. There are two kinds of information:

- **Correct answers**: the number of correct answers you and the other participant each got on the task.
- **Monetary contribution**: the earnings you and the other participant each contributed on the task. This depends on the correct answers and the reward (high or low):
  *correct answers* x *reward per correct answer.*




Two examples of different information orientations. We used all 8 possible orientations of participant and contribution information between subjects, evenly divided across subjects accounting for Advantaged status and Focus treatment. Each subject only saw one orientation to allow them to develop information-seeking patterns.

The information boxes are available for **6 seconds.**

Within this time limit, you can decide which and how many boxes to open. Boxes can be opened more than once.

At times, the program might close some boxes. If this happens, you can't open those boxes again in that round.

You will have the opportunity to practice with these boxes on the next page.

The information boxes are filled with a placeholder number.

In the actual rounds, the information will be based on your performance in Session 1.

This practice will not count toward your bonus.

In the first practice round,
you can familiarize yourself with the **layout** of the information boxes for as long as you want.

In the second practice round,
you will familiarize yourself with the **timing** of the information boxes.

They will be presented for **6 seconds**, like in the actual rounds.

**In summary:**

- In 20 rounds you will divide a common account earned by yourself and another participant.
- Each participants' contribution depends on the correct answers and the reward per correct answer on a single task.
- Before the division, you can inform yourself about correct answers and monetary contributions by hovering your cursor over information boxes.
- Any round could be chosen for payment at the end of the session.

In the first shown slide, the last paragraph "At times, the program might close some boxes" was only included in the constrained Focus treatments and left out in the Free focus treatment.

**Comprehension questions:**

1. On the information screen, what does "correct answers" refer to? MULTIPLE CHOICE [The number of answers you and the other participant each got correct on that task]

2. On the information screen, what does "monetary contribution" refer to? MULTIPLE CHOICE [The earnings (correct answers X reward rate) you and the other participant each contributed to the common account on that task]

**Part 2**

Thank you for completing Part 1 of the Session 2.

Now, we will proceed to Part 2 of this study.

In this part of the study, you will do the same division task as the previous part.

However, you will divide common accounts for **two other, anonymous players** instead:

Player High and Player Low

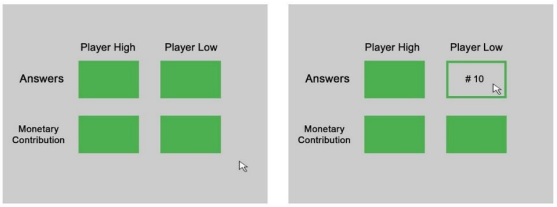
Just like you, both Player High and Player Low have also completed the series of tasks online in Session 1.

Player **High** receives a **HIGH Reward** per correct answer.
Player **Low** receives a **LOW Reward** per correct answer.

As in Part 1, you will be presented with information boxes with Player High's and Player Low's number of correct answers and earnings for **6 seconds**.

And you can **hover the mouse cursor over a box** to see the underlying information.

This is the information layout you will see




In row 2, the right slide switched the information about Players High and Low for Disadvantaged participants so Player Low was described first. The last slide showing the orientation of information varied based on the participant's information orientation. Here, the orientation matched that of Involved trials such that for Advantaged players, Player High's information was in the same row or column as Self information, and Player Low's information was in the same row or column as Self information for Disadvantaged players.

You will complete **20 rounds** of divisions.

For every round, you will make the decision for a different pair of players.

As a reminder:

Your bonus will be based on **randomly selected decision** from **either Part 1 or Part 2**.

If the computer selects one of your decisions from Part 2,

Player High and Player Low from a randomly selected round will be paid according to your decision.

You will receive a fixed bonus of £1.

**Consider each decision carefully** as any of them could be randomly selected for payment towards other participants at the end.

**Comprehension questions:**

1. Which reward condition were Player High and Player Low assigned to? MULTIPLE CHOICE [Player High: High reward, Player Low: Low reward]

2. If Part 2 determines the bonus, how will you be paid? MULTIPLE CHOICE [A set 1 pound bonus]

3. If Part 2 determines the bonus, how will Player High and Player Low be paid? MULTIPLE CHOICE [The amount you gave to each of them on a randomly selected round from Part 2]

**Social Norms**

## Part 3

In Part 3 we will ask you several questions.

At the end of the session, the computer will select one question from Part 3 at random.

In addition to your bonus from Part 1 or Part 2, you can win a bonus depending on your answer to this question.

We will give you more precise instructions about the bonus as you proceed through Part 3

## Part 3.1

In the questions below, please give us your best estimate.

You will earn £1 if you are within 5% of the correct answer.

*Elicitation questions*

1. We selected a random task from Session 1 of the experiment and compared the task performance of 100 members of the HIGH group with the task performance of 100 members of the LOW group. The monetary earnings each person contributed is measured as the number of correct answers in the task times the reward rate. Remember that the reward rate per correct answer was higher in the HIGH group than in the LOW group.

   In how many of these 100 comparisons do you think that the member of the LOW group produced a larger monetary contribution than members of the HIGH group?

2. In Part 1, you were matched with 20 different participants and saw information on both your task performance and the task performance of the matched participants.

In how many of these 20 rounds did the participant you were matched with answered more question correctly than you did?[27]

[27]We asked these two questions only in the Constrained Focus treatments.

**Part 3.2**

In the questions below, *morally appropriate* refers to an action that is "correct", "fair", or "ethical" according to your values and morality.

*Elicitation questions* How did you decide how to split the common account? [OPEN QUESTION]

According to your moral values, how would you judge the following ways of splitting the common account?[28]

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant
3. Splitting the account considering only the number of correct answers of each participant in Session 1
4. Keeping all for oneself

[Possible answers: Very morally inappropriate, Somewhat morally inappropriate, Somewhat morally appropriate, Very morally appropriate]

[28]The order of the norms questions is randomized at the individual level and it it consistent across the different elicitation screen. That is if a participant sees the questions in the order meritocratic, libertarian, egalitarian in the screen about the moral norms, this order is preserved in the following screens as well.

# Part 3.3

You will now have to judge whether the behavior described in some statements is *socially appropriate*.

Those statements are the same you read in the previous webpage.

*Socially appropriate behavior* refers to an action that is "correct", "fair", or "ethical" according to most participants.

**Bonus**

If the computer selects one question from Part 3.3 for payment,

- We will check how participants that split the common account judged the social appropriateness of the behavior described in the question.
- You will win an additional £1 if your judgment coincides with the most common judgment.

**For example, you will see a question like this:**

**1)** How do you judge the following behavior?

"*I split the common account giving to each participant the monetary contribution he/she produced in Session 1.*"

- ○ Very socially inappropriate
- ○ Somewhat socially inappropriate
- ○ Somewhat socially appropriate
- ○ Very socially appropriate

You will win a £1 bonus if you pick the answer that is selected with the highest frequency by the other participants.

Please answer the questions on the next webpage.

*Comprehension questions*

1. For socially appropriate we mean an action that: MULTIPLE CHOICE [Cost people will find "correct", "fair", or "ethical"]

2. If a question from Part 3.2 is selected for payment, you earn a bonus of £BLANK if you: MULTIPLE CHOICE [pick the answer that is selected with the highest frequency by the other participants that divided the common account.]

*Elicitation questions:*

Are the following ways of splitting the common account socially appropriate? Remember to select the answer you think is most common.

1. Giving to each participant the monetary contribution he/she produced in Session 1

2. Giving an equal amount to each participant

3. Splitting the account considering only the number of correct answers of each participant in Session 1

[Possible answers: Very socially inappropriate, Somewhat socially inappropriate, Somewhat socially appropriate, Very socially appropriate]

# Part 3.4

For the next questions you will have to guess how a group of participants judged some behavior.

The groups that you will have to consider are:

- participants that a) received a **low reward** per each correct answer in Session 1 and b) that split the common account in Session 2
- participants that a) received a **high reward** per each correct answer in Session 1 and b) that split the common account in Session 2

**Bonus**

As before, if the computer selects one question from Part 3.4 for payment:

- We will check which is the most common judgment among the group specified by the question.
- You will win an additional £1 if guessed what is the most common answer in that group.

**For example, you will see a question like this:**

**Consider the group of participants that a) received a *low reward* per correct answer in Session 1 and b) split the common account in Session 2**

**1)** How do you think most of participants in this group judged the statement below ?

"*I split the common account giving to each participant the monetary contribution he/she produced in Session 1.*"

- ○ Very socially inappropriate
- ○ Somewhat socially inappropriate
- ○ Somewhat socially appropriate
- ○ Very socially appropriate

You will win a £1 if you guess the most common judgment among the group described in the question.

Please answer the questions on the next webpage.

*Comprehension questions* In Part 3.3 you will have to guess the way most participants in some groups judged a statement. Among the groups below, tick all the ones you will have to consider.

- A group composed of participants that a) received a *low reward* per correct answer in Session 1 and b) split the common account in Session 2 [Correct]
- A group composed of participants that a) received a *high reward* per correct answer in Session 1 and b) split the common account in Session 2 [Correct]
- A group composed of participants that a) received a *low reward* per correct answer in Session 1 and b) *did not* split the common account in Session 2
- A group composed of participants that a) received a *high reward* per correct answer in Session 1 and b) *did not* split the common account in Session 2

*Elicitation questions*

Consider the group of participants that a) received a *HIGH REWARD* per correct answer in Session 1 and b) split the common account in Session 2

How do you think most of participants in this group judged the following ways of splitting the common account?

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant
3. Splitting the account considering only the number of correct answers of each participant in Session 1

Now, consider the group of participants that a) received a *LOW REWARD* per correct answer in Session 1 and b) split the common account in Session 2

How do you think most of participants in this group judged the following ways of splitting the common account?

1. Giving to each participant the monetary contribution he/she produced in Session 1
2. Giving an equal amount to each participant

3. Splitting the account considering only the number of correct answers of each participant in Session 1

*Questionnaire, page 1*

**Please complete the following short survey.**

1. Age: [Open-ended question]
2. Gender: [Possible answer: Man, Woman, Other]
3. What is your nationality? [List of all countries in the World]
4. Generally speaking, where do you place yourself on the left-right political spectrum? [Possible answers: left, center-left, center, center-right, right]
5. How much do you agree with this statement? "The government should take measures to reduce differences in income levels." [Possible answers: Completely disagree, Somewhat disagree, Somewhat agree, Completely agree]
6. What is the highest level of school you have completed or the highest degree you have received? [Possible answers: Less than high school degree, High school degree, Some University but no degree, Bachelor degree, Master degree, Doctoral degree]
7. How much total combined money did all members of your HOUSEHOLD earn last year? [7 different income brackets]

**We ask you the questions below to check whether we need to improve the study. As for every other question in the study, the approval of your submission does not depend on your answers.**

8. Was there anything in the instructions that was unclear? [Open-ended question]
9. How attentive and focused were you in the last rounds of Part 1? [1 to 10 scale. 1 = Not at all attentive or focused, 10 = Completely attentive or focused]
10. How attentive and focused were you in the last rounds of Part 2? [1 to 10 scale. 1 = Not at all attentive or focused, 10 = Completely attentive or focused]
11. What do you think the aim of this study is? [Open-ended question]
12. Do you have any remark or suggestion? [Open-ended question]

*Questionnaire, page 2*

**Please answer the question below.**

In your experience, was/were there any box(es) that was/were more likely to be closed by the program?" [Open-ended question]

*Questionnaire, page 3*

**Please answer the following questions.**

How morally appropriate would you consider the following ways to use the information?

1. using exclusively information about the **participants' monetary contribution** in a task to decide how to split the common account? [1 to 5 scale. 1 = Very inappropriate, 5 = Very appropriate]

2. How morally appropriate would you consider using exclusively information about the **participants' number of correct answers** in a task to decide how to split the common account? [1 to 5 scale. 1 = Very inappropriate, 5 = Very appropriate]