Pre-Analysis Plan

This pre-analysis plan consists of two parts. First, we describe the statistical tests we intend to conduct in the paper. Second, we provide some further theoretical background for our main hypothesis, i.e., that participants with a LOW draw in part 1 (S1<5) are more likely to choose equality in part 2 than those with a high draw (S1>5), and less likely to choose truth-telling.

Preliminaries

1. We define the LOW group and HIGH group as those participants who received a low draw (S1<5) and a high draw (S1>5) in Part 1 respectively. Unless otherwise indicated, all analysis will be carried out only on the participants in these two groups (i.e., we will not use the data from Passive players or those with a random draw of S1=5).
2. As a manipulation test, we run two-sided tests of proportions to check whether HIGH participants were indeed less likely to report equality and more likely to choose truth-telling than high participants. This allows us to see whether our manipulation successfully induced different behavior in the two groups.
3. As a further manipulation test, we will also show regressions of a dummy for choosing a motive on a dummy for the HIGH group and the signal from part 1. We will run a separate regression for both the truth-telling and equality motive. This allows us to see whether the exact draw S1 affected the propensity to select a given motive conditional on being in the HIGH or LOW group.
4. We compute the fraction of participants in each group who chose neither equality nor truth in Part 2. If this fraction is non-negligible (>10%), our main analysis will separately report the results for truth-telling and equality. If not, the results for truth-telling and equality are likely to be identical. Hence we will report only the results for equality in the main text, and only note those cases in which the results for equality and truth-telling differ using e.g., a footnote in the text.

Main Analysis

5. Two-sided test of proportions testing whether the fraction of participants choosing equality in the HIGH group in Part 2 differs significantly from the LOW group. This allows us to test our main hypothesis, i.e., that participant with a HIGH draw in Part 1 are less likely to report equality in Part 2.
6. Two-sided test of proportions testing whether the fraction of participants choosing truth-telling in the HIGH group in Part 2 differs significantly from the LOW group.

This allows us to test our main hypothesis, i.e., that participant with a HIGH draw in Part 1 are more likely to report the truth in Part 2.

Robustness Analysis

7. Linear regression of a dummy for choosing a given motive in Part 2 on a dummy for the HIGH group plus the signal received in Part 1. This allows us to test whether the exact signal received matters even after controlling for whether it was a LOW or HIGH signal. We will run separate regressions for the two main motives (equality and truth-telling).
8. Re-do the main analysis (points 5 and 6) separately for participants who chose the selfish motive in part 1, and those who did not. This allows us to see whether the main effect we observe in points 5 and 6 comes from those who did not choose the selfish motive in part 1, those who did, or both.

Additional Analysis

9. Two-sided Mann-Whitney test investigating whether the difference between the appropriateness scores for equality and truth-telling in part 3 differs between the HIGH and LOW group. In addition, we run a linear regression of the difference between the appropriateness score for equality and the appropriateness score for truth-telling on a dummy for the HIGH group plus the signal received in Part 1. This analysis allows us to test whether members of the two groups differ in the relative appropriateness rankings for the truth-telling and equality motive.

Theoretical Background

Why do we expect individuals in the HIGH group to be more likely to select the truth-telling motive in Part 2, and less likely to select equality? In this section, we will provide a fairly general theoretical framework that provides an intuitive justification for this hypothesis. Specifically, we will consider two very similar frameworks that are informed by two distinct psychological interpretations.

Mental representations framework:

The point of departure for this theoretical framework is the fact that the first party lying/dictator game (LDG) game played in Part 1 is similar to both the classic dictator game (DG), and the Fischbacher and Föllmi-Heusi (2013) lying game (LG) in terms of its basic structure. Therefore, this framework assumes that participants can choose whether to *represent* the LDG in their mind as either the DG (in which they trade off social preferences against self-interest) or the LG (in which they trade of lying aversion motives against self-interest). This implies that if she represents the LDG as the lying game, she chooses the same action as she would if she was playing the lying game, and if she represents the game as a dictator game, she chooses the same action she would when playing the dictator game.

We model this by assuming that each individuals has a probability, $\delta(.)$, of representing the game as lying game, and otherwise represents the game as a dictator game. Crucially, we also hypothesize that the $\delta(.)$ will differ systematically as a result of the random number drawn by the subject. Specifically, we hypothesize that participants conveniently choose to represent the LDG as the game (either LG or DG) that allows them to achieve the highest personal payoff given the random number drawn. Consider, for example, an individual with a low draw (S1<5). Representing the LDG as an LG may force this individual to truthfully report her draw if her lying costs are sufficiently large. By contrast, representing the LDG a DG would assure the individual of a payment of at least 5 (equal payoffs for both individuals), provided that the individual does not want to give away more than half the endowment to the other participant in the DG. As a result, this individual may be able to achieve a higher personal payoff given her low draw if she

conveniently chooses to represent the LDG as a DG. Conversely, individuals with a high draw may benefit from representing the LDG as a LG.

A basic theoretical framework that serves to organize some of these intuitions is provided by the following:

$$U(s,r) = \delta(s) * R^{LG}(s,r) + [1 - \delta(s)] * R^{DG}(r)$$

where s is the random number drawn and r is the report made by the individual. $R^{LG}(s,r)$ refers to the utility the participant would obtain if she represented the LDG in her mind as identical to the LG, while $R^{DG}(r)$ refers to how the utility the participant would obtain if her mental representation of the LDG were identical to the DG.

The probability weights placed on each representation are reflected by $\delta(s)$. Crucially, we hypothesize that participants will engage in motivated reasoning and shift the probability that they interpret the LDG as either the LG or the DG in a self-serving way that depends on their random number draw, s. More specifically, we conjecture that for $s' > s$ we will have $1 \geq \delta(s') \geq \delta(s) \geq 0$, with $\delta(s') > \delta(s)$ for at least one $s' > s$. If correct, this implies that individuals with low number draws are more likely to represent the game as a DG and average behavior will be closer to behavior in the DG, while individuals with higher random number draws will be more likely to represent the game as an LG and average behavior will be closer to that observed in the LG. One can think of this as reflecting situations in which individuals who end up in different roles (e.g. rich vs poor, lucky vs unlucky) form different mental representations of the situation in a way that suits their own purposes.

We then assume that, once the individual has decided to interpret the LDG as either LG or DG in part 1 of the experiment, she will use the same representation in part 2. Since part 2 consists of a third-party LDG where the individual herself has no stake in the outcome, this implies that the individual will choose to equalize payoffs if she chose to represent the LDG in part 1 as a DG, and will report the random draw if she chose to represent the LDG as an LG in part 1. If individuals with low (high) draws are indeed more likely to choose equality (truth-telling) in part 1, we then expect the random draw in part 1 to have a similar effect on the motive selected in part 2 (i.e., δ will be identical

in both parts). Note that in our analysis we allow both for a discrete jump in δ at S1=5 (points 5 and 6), and for δ to increase linearly in s (point 7).

Motives Framework:

The second theoretical framework is similar, but has a slightly different interpretation. In particular, rather than assuming that an individual decision maker represents the LDG as either the LG or the DG when making her report decision, the motives framework effectively assumes that the individual obtains weighted average of the utility received in each game. In other words, the weighting over games now takes place within an individual. However, again, these weights are formed in a motivated way as a function of the random number drawn, implying similar comparative static predictions.

The point of departure for this interpretation of the model takes the perspective that individuals face different *motives*, and therefore have to choose how much weight to place on each of the motives which may play a role in a given context. In the context of our first party LDG game, the participants face three primary motives: (i) a preference for truth-telling (or lying aversion), (ii) social preferences, and (iii) self-interest. In the formulation below, the individual always has a constant weight on self-interest, but is able to shift how much weight she places on her preference for truth-telling vis-à-vis her social preferences. As in the representations framework discussed previously, we conjecture that the relative weight that the individual places on her two "moral motives" will vary in a self-interested way according to the random draw that she receives. In particular, upon receipt of a low random draw, a participant may inflate the weight given to her social preferences (e.g. inequity aversion) in order to justify lying, while a participant who receives a high random draw may inflate her lying aversion / truth-telling motive and use it as an *excuse* in order to justify choosing an inequitable split.

If one writes $R^{LG}(s,r) = r + M^{T}(s,r)$ and $R^{DG}(r) = r + M^{S}(r)$, we can transform the representations framework into the motives framework as follows:

$$U(s,r) = r + \delta(s) * M^{T}(s,r) + [1 - \delta(s)] * M^{S}(r)$$

where, again, s is the random number drawn and r is the report made by the individual and therefore also reflects the money she receives (her monetary self-interest). $M^T(r)$ reflects that individual's truth-telling motive, while $M^S(r)$ reflects the individual's social preferences. For example, $M^T(r)$ may be the Abeler et al. (2018) model of lying costs, whereas $M^S(r)$ may be a standard social preference model such as Fehr and Schmidt (1999). In this framework, we interpret $\delta(s)$ as the within-individual weight placed on the truth-telling motive. Regardless of the specific model, as before we conjecture that for $s' > s$ we will have $1 \geq \delta(s') \geq \delta(s) \geq 0$, with $\delta(s') > \delta(s)$ for at least one $s' > s$.

Similar to the representations framework, we then once again assume that the weight $\delta(s)$ put on truth-telling in part 1 will carry over to part 2. Since part 2 consists of a third-party LDG where the individual herself has no stake in the outcome, this implies that the individual will choose to equalize payoffs if her $\delta(s)$ is sufficiently low, and will report the random draw if her $\delta(s)$ is sufficiently high. If individuals with low (high) draws indeed have a higher δ in part 1, we then expect individuals with a low (high) draw to also be more likely to report equality (the truth) in part 2. Note that in our analysis we allow both for a discrete jump in δ at S1=5 (points 5 and 6), and for δ to increase linearly in s (point 7).

- **References**

Abeler, Johannes, Daniele Nosenzo, and Collin Raymond. Forthcoming. "Preferences for truth-telling." *Econometrica*.

Fehr, Ernst, and Klaus M. Schmidt. 1999. "A theory of fairness, competition, and cooperation." *The Quarterly Journal of Economics,* 114(3): 817-868.

Fischbacher, Urs and Franziska Föllmi-Heusi. 2013. ”Lies in Disguise—An Experimental Study on Cheating.” *Journal of the European Economic Association*, 11(3): 525-547.