Motivated Beliefs about Others' Preferences

Pre-analysis Plan | June 2026

Shuran Gao

University College London

Research Question

In many real-world settings, beliefs about others' preferences influence decisions that involve self-interest. In such environments, individuals may form motivated beliefs: beliefs shaped not only by available evidence but also by what the individual would like to be true. Motivated beliefs can serve a range of motives, from protecting one's self-image to sustaining optimism about the future. This study focuses on the instrumental motive, in which a belief is convenient because it rationalises a self-interested action. We examine whether individuals form motivated beliefs about others' preferences when such beliefs can rationalise self-interested behaviour. In particular, we test whether individuals report higher beliefs about another person's willingness to work (WTW) when doing so supports assigning more tasks to that person.

This study contributes in two main ways. First, it studies motivated beliefs in an applied employer–worker setting, in which one party (the Employer) makes a payoff-relevant assignment decision while forming beliefs about the other party (the Worker). Most evidence on motivated beliefs concerns beliefs about oneself, such as one's own ability, health, or future prospects; far less is known about whether the same distortions arise in beliefs about others, even though such beliefs drive many consequential economic decisions, including how managers allocate work, how negotiators read a

counterpart, and how people judge what others are owed. Although the setting is deliberately artificial, it isolates the central feature of these interactions: a belief about someone else that can be bent to rationalise a self-interested choice. Second, we investigate higher-order beliefs to examine whether individuals expect others to form motivated beliefs, and whether those who form motivated beliefs are aware of their own motivated reasoning.

Experimental Design

Overview

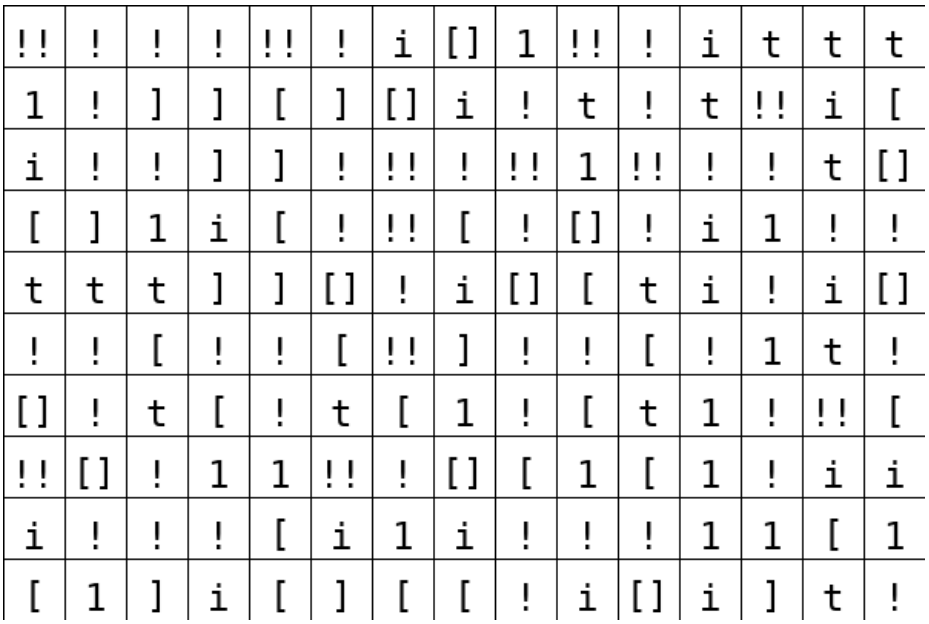
To address our research question, we design an online experiment that creates conditions in which individuals may have an incentive to distort their beliefs about others' preferences due to the presence of self-interest, as well as conditions in which such motives are absent. The experiment will be conducted on the Prolific platform with a target sample of 800 adult participants located in the United States, recruited before any exclusions. We prescreen participants with a history of completing 100+ previous studies and maintaining a 95% to 100% approval rate. Participants who fail an attention check are excluded from the analysis sample. The central theme of the experiment is participants' WTW on a set of real-effort counting tasks. Participants report their own WTW and make predictions about another Worker's WTW.

Participants are randomly assigned to one of four groups defined by role (Employer or Worker) and incentive condition (Low or High). Employers make a payoff-relevant assignment decision, while Workers act as pure predictors; the incentive condition varies the bonus Employers receive from assigning tasks, generating variation in the strength of self-interest at stake. If motivated beliefs play a role in this environment, stronger incentives should create a stronger motive to rationalise higher task assignments. The between-subjects variation in incentives therefore allows us to assess the robustness of the proposed motivated reasoning mechanism.

Willingness to Work

The real-effort tasks require participants to count the number of occurrences of a randomly selected symbol within a grid of symbols and enter the correct number to pro-

ceed. These tasks are designed to be mechanical, unfamiliar and moderately tedious in order to generate disutility of effort while minimising the role of prior experience or task-specific ability. An example of such counting task is provided below.



WTW is defined as the maximum number of tasks a participant is willing to complete in exchange for a fixed payment. We incentivise elicitation of WTW using a Becker-DeGroot-Marschak (BDM) mechanism. Participants first choose a number between 0 and 25 representing the maximum number of tasks they are willing to complete for a fixed payment. After the choice is made, a random integer between 0 and 25 is drawn by the system. If the randomly drawn number is less than or equal to the participant's stated WTW, the participant completes that number of tasks and receives the payment. Otherwise, the participant is not required to complete any tasks and receives no payment from the task stage.

Treatment Arms

After the WTW elicitation, participants will be randomly assigned to one of four groups defined by role and incentive condition: Low-Incentive Employer, High-Incentive Employer, Low-Incentive Worker, and High-Incentive Worker.

Participants assigned to the role of Employer make two main decisions. First, they choose how many tasks, from 0 to 25, to assign to the matched Worker. Second, they

report their belief about their matched Worker's WTW. Employers receive a bonus that increases with the number of tasks assigned. The incentive condition varies the size of this bonus: Employers receive 5 cents per task assigned in the low incentive condition and 25 cents per task assigned in the high incentive condition. Employers receive this bonus based on their assignment decision regardless of whether the Worker completes the assigned tasks, which removes any strategic concerns related to task completion. In contrast, Workers do not make assignment decisions. Instead they are pure predictors and report beliefs about another Worker's WTW.

Prediction Questions

The prediction stage elicits the beliefs listed below. All predictions are incentivised using a binarised scoring rule (BSR) under a pay-one scheme, in which one prediction is randomly selected for payment.

(P1) First-order belief about a Worker's WTW.
The number of tasks the participant believes a Worker is willing to complete for the stated payment. Employers predict their matched Worker; Workers, as pure predictors, predict another Worker. Elicited from both roles.

(P2) First-order belief about task assignment.
The number of tasks the Worker believes the Employer will assign. Elicited from Workers.

(P3) Employer second-order belief.
The Employer's belief about how another Worker, acting as a pure predictor, would predict the matched Worker's WTW.

(P4) Worker second-order belief about the Employer.
The Worker's belief about their matched Employer's first-order belief of the Worker's own WTW.

(P5) Worker second-order belief about another Worker.
The Worker's belief about how another Worker, acting as a pure predictor, would predict the Worker's own WTW.

Outcomes and Estimation Strategy

Primary Outcome

Our primary outcome is the difference in first-order beliefs about a Worker's WTW (P1) between Employers (treatment) and Workers (control). Both groups predict the WTW of a Worker drawn from the same population, so the prediction task is identical across groups; they differ only in that Employers hold a payoff-relevant stake in the assignment, whereas Workers are pure predictors. Absent motivated reasoning, the two groups should predict the same on average, so a positive difference identifies motivated belief. We measure the belief as the number of tasks the participant expects the Worker to be willing to complete for the stated payment.

Hypothesis 1 *Employers on average make a higher prediction than Workers about a Worker's WTW.*

To test Hypothesis 1, we estimate the following OLS regression:

$$Y_i = \alpha + \beta \times treat_i + \gamma \times X_i + \delta \times Z_i + \varepsilon_i,$$

where Y_i is the participant's belief (P1) about a Worker's WTW, $treat_i$ is a dummy variable equal to one if participant i is allocated to the Employer role, X_i controls for the participant's own WTW, and Z_i denotes demographic controls (age, and dummies for education, gender and income categories). The coefficient of interest is β, which captures the average treatment effect on beliefs about a Worker's WTW. Throughout, we report heteroskedasticity-robust standard errors.

Secondary Outcomes

Incentive manipulation and the motivated-reasoning mechanism. We analyse two outcomes on the Employer sample, both estimated with a common specification,

$$Y_i = \alpha + \beta \times high\_incentive_i + \gamma \times X_i + \delta \times Z_i + \varepsilon_i,$$

where $high\_incentive_i$ is a dummy equal to one if participant i is in the high-incentive condition, X_i controls for the participant's own WTW, and Z_i denotes demographic controls.

Hypothesis 2 *High-Incentive Employers on average assign a higher number of tasks than Low-Incentive Employers.*

Hypothesis 3 *High-Incentive Employers on average make a higher prediction about a Worker's WTW than Low-Incentive Employers.*

Hypothesis 2 serves as a validation check for the incentive manipulation, and Hypothesis 3 follows from the proposed motivated reasoning mechanism: if stronger assignment incentives create a stronger motive to justify higher task assignments, High-Incentive Employers may both assign more tasks and report higher beliefs about Workers' willingness to work. For Hypothesis 2, Y_i is the number of tasks assigned by participant i; for Hypothesis 3, Y_i is the Employer's belief (P1) about the matched Worker's WTW. In each case the coefficient of interest is β.

Higher-order beliefs. We treat higher-order beliefs as a central secondary outcome, though we do not commit to directional predictions for them. They allow us to examine whether individuals anticipate motivated reasoning in others, and whether motivated reasoners are aware of their own distortion. Each outcome is a within-subject difference in beliefs. For each gap G_i we estimate

$$G_i = \alpha + \beta \times high\_incentive_i + \varepsilon_i,$$

where the intercept α is the average gap in the low-incentive condition and β is the difference in the average gap between incentive conditions. The two gaps are estimated on different samples.

To examine awareness of one's own motivated belief, we use the within-Employer gap $G_i = P1_i - P3_i$, the difference between the Employer's first-order belief (P1) and second-order belief (P3) about the same matched Worker, estimated on the Employer sample. A positive α would be consistent with Employers having partial awareness of their own motivated belief, and β measures whether this gap varies with incentive strength.

To examine awareness of others' motivated beliefs, we use the within-Worker gap $G_i = P4_i - P5_i$, the difference between the Worker's second-order belief about the matched Employer (P4) and about another Worker acting as a pure predictor (P5), estimated on the Worker sample. Since Workers are informed of their matched Employer's

incentive condition, a positive α would be consistent with Workers anticipating that Employers form motivated beliefs, and a positive β would indicate that they anticipate stronger motivated beliefs from Employers facing stronger incentives.

Exploratory Analyses

In addition to the primary and secondary outcomes, we will conduct several exploratory analyses related to prediction accuracy, the anticipation of assignment, information use, and treatment-effect heterogeneity in this environment. These analyses are not tied to directional predictions and will be interpreted as exploratory rather than confirmatory. Where applicable, we estimate them using the same OLS framework as above, controlling for own WTW and demographics.

First, we revisit interpersonal projection, the well-documented pattern in which people's predictions of others' preferences are pulled towards their own. As a benchmark, we test whether Workers, who act as pure predictors with no stake in the assignment, display this pattern, namely whether their predictions of a Worker's WTW increase with their own WTW. Confirming projection among pure predictors would replicate prior findings and establish a baseline against which the beliefs of self-interested Employers can be compared. We then test whether Employers predict less accurately than Workers, and whether higher incentives are associated with larger departures from the target Worker's actual WTW.

Second, we will analyse Workers' predictions about the number of tasks Employers will assign (P2), examining whether they anticipate the assignment behaviour generated by the incentive manipulation, and in particular whether they expect higher-incentive Employers to assign more tasks. Since the assignment is the behavioural counterpart of the Employer's motivated belief, such evidence would complement the higher-order belief analysis on whether Workers anticipate others' motivated reasoning.

Third, towards the end of the survey we offer Employers the option to revise their assignment in light of the Worker's actual WTW. The revised assignment replaces the original and determines the bonus accordingly. Employers who revise can condition their assignment on this information through a threshold rule, assigning a chosen number of tasks depending on whether the Worker's WTW falls above or below a stated cut-off. The decision of whether to revise is itself informative: an Employer who has formed a motivated belief rationalising a high assignment may prefer not to incorpo-

rate the Worker's true WTW, since it could undermine that justification. We will examine whether the propensity to revise is lower among high-incentive Employers and how it relates to the size of the initial assignment. Reluctance to revise would be consistent with a motive to protect, rather than correct, a self-serving belief.

Finally, we will explore heterogeneity in the main effects across participant characteristics. Using interactions with the treatment and incentive indicators, we will examine whether the effects on beliefs differ by demographics such as gender, age, education, and income, and by participants' own WTW. We will also check that the main results are robust to including these characteristics as controls.