

Replication of “Does Identity Affect Labor Supply?” (Oh, 2023)

Development Economics

SHARMA Vivan, PEREIRA Rodrigo

March 17, 2025



SciencesPo.

Department of Economics
Sciences Po, Paris

1 Summary

Theories of identity suggest that individuals in the labor market may avoid attractive job opportunities if they clash with their sense of identity. If people refrain from pursuing certain careers despite having the necessary aptitude due to identity concerns, it could lead to an inefficient allocation of talent within the economy. However, there is limited empirical evidence on the extent and mechanisms through which identity concerns influence labor market decisions. Establishing the impact of identity using observational or survey data alone is challenging, and experimentally isolating its role is equally difficult. "Does Identity Affect Labor Supply?", by Oh (2023), addresses these challenges by integrating insights from social psychology and leveraging unique aspects of the Indian caste system to experimentally examine how identity influences job-specific labor supply. The findings reveal that casual workers in Odisha, India, are averse to taking on jobs associated with castes other than their own, especially when those castes rank lower in the social hierarchy. These effects persist even when job decisions are made in private, with no external visibility of job engagement or performance, underscoring the importance of internal identity concerns. A supplementary experiment further quantifies the wage penalty workers are willing to incur to avoid tasks associated with other castes, showing that a significant proportion of workers are willing to forgo up to ten times their daily wage to avoid spending as little as ten minutes on such tasks.

The implications for labor markets are ambiguous. Workers may forgo high-paying jobs with desirable working conditions due to concerns about preserving their caste identity. Fears of social sanctions or a preference for occupations aligned with their caste networks can further reinforce the connection between identity and employment choices. Labor specialization can improve productivity by encouraging individuals to pursue the same occupations as their parents or network members, but Indian firms may need to hire multiple workers instead of one for the same maintenance jobs as compared to those in other settings.

In the following sections, we discuss the theoretical setup and replicated outputs of the paper - their estimating equations, underlying assumptions, and interpretations.

2 Setup

The conceptual framework about a worker's decision problem of job take-up used to motivate the experimental design is presented below.

A worker ' i ' of caste ' c_i ' expects to spend time ' t_0 ' working on the default task and ' t_k ' on the extra task ' k '. Total time worked is therefore $T = t_0 + t_k$. ' M_i ' and ' L_i ' indicate worker i 's utility for money and leisure respectively. A representative utility function looks like:

$$U_{ik}(c_i, t_k) = \underbrace{M_i}_{\text{Money}} + \underbrace{L_i}_{\text{Leisure}} - \sum_{g \in \{0, k\}} \left[\underbrace{V_{ig}(c_i, t_g)}_{\text{Variable Cost}} + 1\{t_g > 0\} \cdot \underbrace{F_{ig}(c_i)}_{\text{Fixed Cost}} \right]$$

When deciding whether to take-up a task, workers must ponder the benefits and the fixed ($F_{ig}(c_i)$) and variable ($V_{ig}(c_i, t_g)$) (time-dependent) costs. Identity concerns are assumed to be a fixed cost. **The first assumption made is that the variable cost is zero if the task is not taken-up, and continuous in time spent on the task.** Given this assumption, fixed costs are estimated by observing worker reactions as the time requirement on the task approaches zero. However, these fixed costs may still include costs other than identity concerns - such as learning, if they are performing the task for the first time. Thus, fixed costs of working on task k are defined as:

$$F_{ik}(c_i) = f_k + \beta_{ik} I_k(c_i)$$

Where f_k is the "inherent" fixed cost, not depending on identity. $I_k(c_i)$ is an indicator for task ' k ' being associated with a caste other than ' c_i '. The coefficient β_{ik} captures the cost of identity violations. **An assumption made here is that a worker does not obtain additional utility from performing a task congruent with their own identity.** This is important to interpret the results obtained later, as it is an alternative way to explain the increased likelihood of workers to reject identity-task offers.

In order to establish the existence of identity concerns, the author uses control tasks and several different caste groups to make a difference-in-difference (DiD) comparison. They find tasks which are very similar to identity

tasks but that are not associated to any particular caste (control tasks). They compare the difference in acceptance rates between identity-related tasks that do not match the worker’s social group and identity-related tasks that do. They then compare this difference to the corresponding difference for non-identity-related tasks. This allows them to experimentally isolate the effect of identity concerns on task acceptance.

Let there be two large groups of workers belonging to categories A and B, evaluating job offers of extra tasks b (associated with group B and therefore can cause identity-related concerns in A workers) and u (no association). $\theta_i = U_{ik}(c_i, 0)$ represents the net utility from taking up the offer that only involves the default task. **A key assumption made here is that the distributions of f_b , f_u , and θ_i satisfy:**

$$\Pr(f_b > \theta_i \mid i \in A) - \Pr(f_u > \theta_i \mid i \in A) \leq \Pr(f_b > \theta_i \mid i \in B) - \Pr(f_u > \theta_i \mid i \in B)$$

The assumption is that the difference in the probability of rejecting a task due to inherent fixed costs is smaller for identity-related tasks than for control tasks across the two groups. This means that when interpreting the difference-in-differences, if the difference in take up is larger for the identity-related task than for the control task, it must be due to identity concerns, given that we have separated fixed costs into inherent and identity-related components. Also, even if an identity-inconsistent task and its paired control task (one that is free of caste associations but involves similar variable cost of effort) involve slightly different inherent fixed costs, the difference in these costs could be similar across workers in both groups. Therefore, if this assumption holds, any significant difference-in-difference estimator in the comparison of take-up rates can be attributed solely to identity concerns. They extend this analysis with multiple social groups to make a more specific prediction - the share of workers with identity concerns would be larger when the task is associated with a lower-status group.

3 Table 2

For this table, the author uses the following DiD specification:

$$Y_{ikt} = \sigma^d \text{different}_{ik} + \lambda^d \text{identity}_k \cdot \text{different}_{ik} + \sigma^l \text{lower}_{ik} + \lambda^l \text{identity}_k \cdot \text{lower}_{ik} + \tau Z_t + P'_k \rho + X'_i \nu + \epsilon_{ikt} \quad (1)$$

Y_{ikt} is an indicator of whether worker i is willing to accept a job offer that involves some time indexed by t on task k . The independent variables include indicators for whether task k is an identity task, a different task, and/or a lower task (a task associated with a lower caste), as defined in the Task Survey. P_k is a vector of task-specific dummies, and X_i is a vector of worker or caste-level dummies. Z_t accounts for potential time-related factors that might influence the likelihood of accepting a job. Standard errors are clustered at the worker level.

Table 2: Identity Inconsistency and Job Offer Take-Up

	Willing to take up job offer				
	(1)	(2)	(3)	(4)	(5)
<i>Identity × Different</i>	-0.233*** (0.043)	-0.233*** (0.043)	-0.231*** (0.045)	-0.244*** (0.046)	-0.243*** (0.047)
<i>Identity × Lower</i>	-0.238*** (0.031)	-0.238*** (0.031)	-0.246*** (0.033)	-0.205*** (0.036)	-0.206*** (0.036)
<i>Different tasks</i>	-0.069* (0.028)	-0.069* (0.028)	-0.060* (0.029)	-0.054 (0.031)	-0.058 (0.031)
<i>Lower tasks</i>	0.129*** (0.028)	0.129*** (0.028)	0.128*** (0.030)	0.111*** (0.032)	0.116*** (0.032)
<i>Hours on extra tasks</i>	-0.055*** (0.005)	-0.055*** (0.005)	-0.053*** (0.006)	-0.053*** (0.006)	-0.053*** (0.006)
Sample Size	15120	15120	13224	13224	13224
R^2	0.23	0.54	0.54	0.55	0.55
Adjusted R^2	0.23	0.52	0.52	0.53	0.53
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$					

λ^d estimates how take-up changes when offers involve a different identity task relative to a different task with no identity connotations (paired control task - as explained above), and λ^l estimates how take-up changes when task k involves a lower identity task relative to a lower paired control task.

Columns 1 & 2 differ only in fixed effects (caste vs. worker). Controlling for worker fixed effects ensures that any individual characteristics affecting workers' willingness to engage in jobs do not change the estimates. Column 3 uses a filtered dataset (when survey data was available). Columns 4-5 add both linear and binary demographic controls, for age, education and wealth specifically. The results are very similar across columns. Same-ranked tasks have similar take-up rates for both identity and control tasks, at 72% (means of same-ranked tasks). Take-up decreases by 23 percentage points if the extra task is an identity task with associations of a different caste, and by another 24 p.p. if it is associated to a lower caste in particular. To put this into perspective, the impact on the take-up rate of spending one more hour on extra tasks is a decrease of only 5.5 percentage points.

4 Table 3

For this table, the author uses the following specification:

$$Y_{ikt} = \lambda^d \text{identity}_k \cdot \text{different}_{ik} + \lambda^l \text{identity}_k \cdot \text{lower}_{ik} + \lambda^{pd} \text{public} \cdot \text{identity}_k \cdot \text{lower}_{ik} + \lambda^{pl} \text{public} \cdot \text{identity}_k \cdot \text{different}_{ik} \\ + \sigma^d \text{different}_{ik} + \sigma^l \text{lower}_{ik} + \sigma^{pd} \text{public} \cdot \text{different}_{ik} + \sigma^{pl} \text{public} \cdot \text{lower}_{ik} + \lambda^p \text{public} \cdot \text{identity}_k + \sigma^p \text{public} \\ + \tau Z_t + P'_k \rho + X'_i \nu + \epsilon_{ikt} \quad (2)$$

As pointed out before, social image and fear of reprisals may influence the decision to take up a job. This would explain the correlations found before through a mechanism other than identity concerns. To test this hypothesis, the author randomizes workers into private or public decision-making groups at the village level and runs the main regressions with the key covariates interacting with the public treatment dummy. Standard errors are clustered at the worker level.

Table 3: Role of Social Image Concerns

	Willing to take up job offer				
	(1)	(2)	(3)	(4)	(5)
<i>Identity</i> \times <i>Different</i>	-0.223*** (0.058)	-0.223*** (0.058)	-0.197** (0.060)	-0.199*** (0.060)	-0.200*** (0.060)
<i>Identity</i> \times <i>Lower</i>	-0.253*** (0.040)	-0.253*** (0.040)	-0.273*** (0.043)	-0.234*** (0.045)	-0.234*** (0.045)
<i>Public</i> \times <i>Identity</i> \times <i>Different</i>	-0.019 (0.084)	-0.019 (0.084)	-0.066 (0.088)	-0.087 (0.086)	-0.083 (0.087)
<i>Public</i> \times <i>Identity</i> \times <i>Lower</i>	0.030 (0.052)	0.030 (0.052)	0.051 (0.055)	0.056 (0.054)	0.053 (0.054)
<i>Hours on extra tasks</i>	-0.055*** (0.005)	-0.055*** (0.005)	-0.053*** (0.006)	-0.053*** (0.006)	-0.053*** (0.006)
Sample Size	15120	15120	13224	13224	13224
R^2	0.23	0.54	0.54	0.55	0.55
Adjusted R^2	0.23	0.52	0.52	0.53	0.53
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$					

From the table, we can see that taking the decision in public has a negligible and not statistically significant effect on take-up. The coefficients on the other variables of interest remain the same as in Table 2. The results are similar across columns, with the differentiation in specifications being the same as in Table 2. Column 2 uses worker fixed effects, Column 3 uses only the data for which survey information was available, and 4-5 add linear and binary demographic controls. These results imply that declining offers with identity-violating tasks is primarily driven by innate identity conflicts and not social image.

5 Figure 3

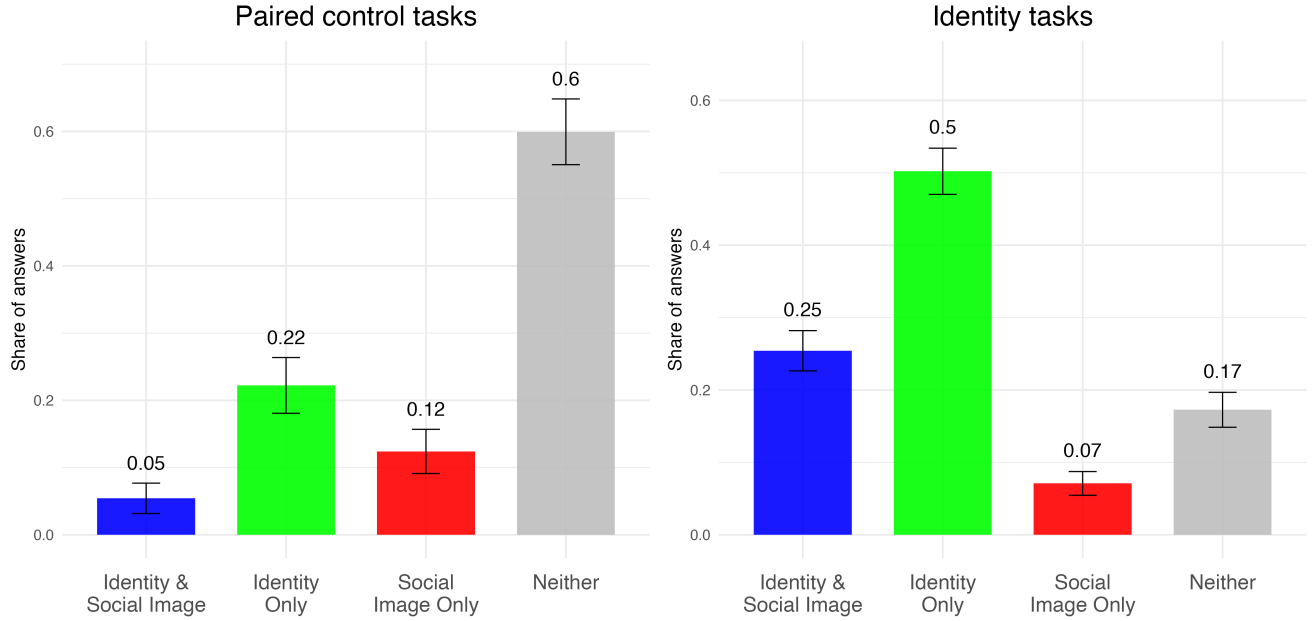


Figure 3: Reasons for Turning Down Job Offers

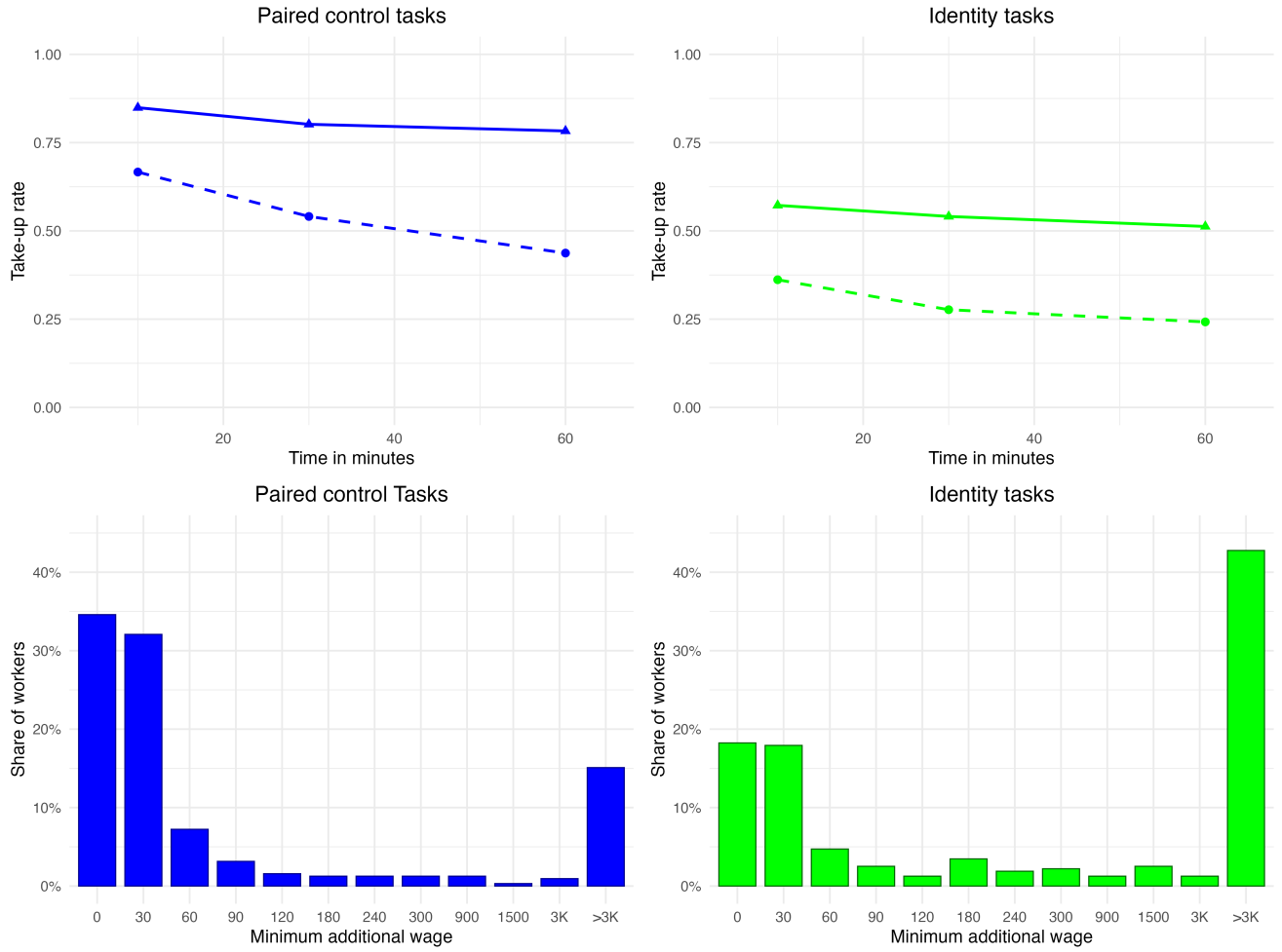
As seen in Table 3, taking decisions in public or private does not seem to have a significant impact on take-up. However, this may be due to the relationship between identity and social concerns. In other words, if a worker has identity concerns then one may argue that he is very likely to have social concerns as well. Conversely, if a worker does not have identity concerns, then it is unlikely that he has social concerns. If this is the case, those who refuse the job in public would take the same action in private. To assess this hypothesis, workers were asked in the follow-up survey why they turned down jobs. Figure 3 illustrates the answers. For identity tasks, 50% of the sample state that identity concerns only were the reason for turning down job offers, and 25% considered both identity and social image concerns. However, only 7% justified their decision on social image grounds only. This is consistent with the hypothesis that both identity and social image concerns impact the decision to accept a job. On the other hand, the majority of workers who turned down jobs with paired control tasks cited neither identity issues, social image or both as their primary cause.

6 Figure 4

To conclude the paper, the author estimates how much money workers are willing to forgo not to take up an identity task. In a similar experiment, 106 workers were offered to switch from a default task to another task for part of the remaining time. This new task randomly varied in type (identity or paired control) and time. These offers involved a bonus payment on top of their daily wage of 300 rupees, itself varying from 0 to 3000 rupees.

Figure 4 presents the results for this experiment. Panel A plots take up rates as a function of time spent in the extra task. Dashed lines plot take up rates when bonus is 10% of daily wage and solid lines correspond to take up when offer's bonus is 10 times the daily wage. The slopes are negative, implying greater variable costs than those for the default task. As expected, take up rates are much lower for identity tasks than for paired control tasks - 43% of workers refuse to spend as little as 10 minutes on identity tasks, whereas only 15% do for paired control tasks. The bonus for these tasks was ten times the daily wage, which showcases just how averse workers are to engaging in identity tasks.

Panel B of Figure 4 shows the minimum amount of bonus workers demand for accepting identity versus paired control tasks. It is a bimodal distribution, with workers either demanding very low or very high bonuses. Identity



7 Table 4

For the final table, the authors use the following specification:

$$refuse_all = \lambda^l identity + \lambda^{pl} public \cdot identity + \tau Z_t + \varepsilon \quad (3)$$

The dependent variable is a dummy reflecting the decision of a worker to refuse all offers involving a task regardless of time and bonus amounts. The independent variables are dummies for identity tasks being offered as well as publicizing their decisions. This analysis is part of the supplementary experiment explained in Figure 4.

Table 4: Caste Inconsistency and Refusal of All Offers Involving a Task

	Refuse all offers regardless of bonus				
	(1)	(2)	(3)	(4)	(5)
<i>Identity tasks</i>	0.277*** (0.033)	0.277*** (0.033)	0.313*** (0.060)	0.272*** (0.048)	0.318*** (0.069)
<i>Public × Identity</i>				0.010 (0.067)	-0.008 (0.066)
Sample Size	636	636	630	636	630
R^2	0.14	0.63	0.65	0.63	0.65
Adjusted R^2	0.14	0.55	0.56	0.55	0.56
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$					

As expected from Figure 4, Panel A, the OLS regression shows workers are 28 p.p. more likely to refuse identity tasks than paired control tasks, controlling for task and caste/worker fixed effects. Results vary slightly across columns. Column 1 has caste fixed effects, while column 2 has worker level fixed effects. Column 3 adds demographic controls, while columns 4-5 add the interaction term of public decision making and identity tasks being offered, as well as the same demographic controls as column 3 respectively. Like we saw earlier, the decision being public or private has very little effect, as seen in columns 4 and 5.

8 Replication Comments

We were able to replicate all of the chosen output to almost perfect degrees. There are slight differences in the standard errors (< 0.01) of some of the coefficients in the regression tables. These are due to computational differences in variance estimations across the two softwares (STATA - original language, and R). We found one small error in the ordering of code in one of the cleaning files (for wealth PCA, commented in the code), but it did not impact the results of the analysis.