Supplemental Appendix for Online Publication

Peers and Motivation at Work: Evidence from a Firm Experiment in Malawi

Lasse Brune
Eric Chyn
Jason Kerwin

May 7, 2020

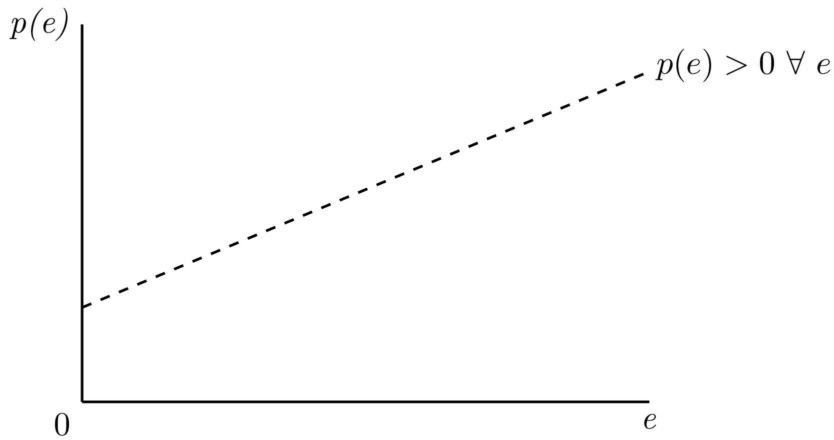
A Appendix Figures and Tables

Figure A1: Photograph of a Tea Field at Lujeri Tea Estates

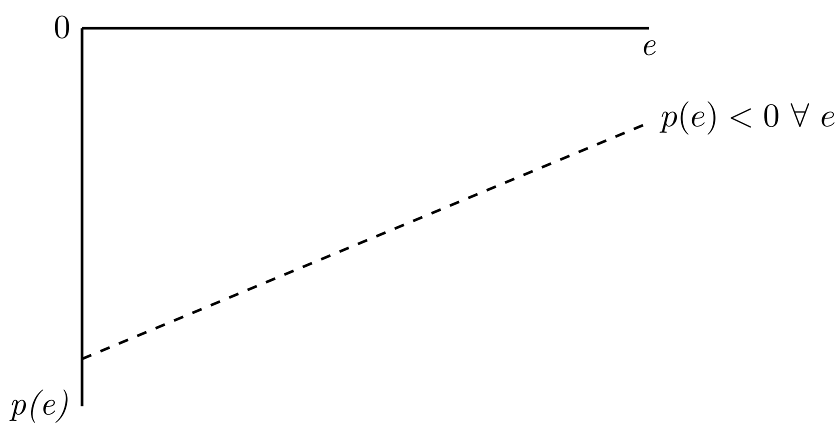
Notes: This photograph shows a tea field at Lujeri Tea Estates. The photograph was taken by the authors in November 2014.

Figure A2: Illustration of Peer Pressure Function Cases

(a) High-Ability Peers Increase Utility



(b) High-Ability Peers Reduce Utility



Notes: This figure illustrates two possible cases for the peer pressure function $p(.)$ from the stylized model of utility presented in Section 7. The x-axis shows effort (e). Panel A shows shows the case where having high ability peers increases total utility. This is consistent with the idea that high-ability peers provide motivation. Panel B shows the case where having high ability peers reduces utility. This is consistent with the idea that high-ability peers reduce utility due to shame or rank preferences.

Table A1: Regression of Worker Ability on Worker Attributes

	Dependent Variable: Worker Ability	
	(1)	(2)
Female	-5.676***	-6.088***
	(2.062)	(2.080)
Married	2.127	2.214
	(2.096)	(2.116)
Household Size	0.852	0.701
	(0.519)	(0.524)
Household Spending per Capita	0.000	0.000
	(0.000)	(0.000)
Age	-0.015	
	(0.068)	
Quartiles of Age		
Age 30 to 35		2.113
		(1.647)
Age 36 to 44		2.049
		(1.801)
Age 44 to 72		1.766
		(1.877)
Experience	0.272***	
	(0.090)	
Quartiles of Experience		
2.1 to 5 Years		5.694***
		(1.671)
5.1 to 10.7 Years		6.258***
		(1.645)
10.8 to 49.5 Years		7.742***
		(1.862)
Worker Fixed Effects	Yes	Yes
Date by Location Fixed Effects	Yes	Yes
Observations	909	909
Adjusted R-squared	0.062	0.075

Notes: This table presents results from a regression of workers' ability levels, as measured in predicted kilograms of tea plucked per day, on various exogenous covariates. The underlying dataset is a cross-section at the worker level. Heteroskedasticity-robust standard errors in parentheses.

Table A2: Tests of Working Sorting Based on Social Networks

	Dependent Variable: Any Neighbors are Friends (=1)	
	(1)	(2)
Cycle Day 1 (no random assignment)	0.082***	0.081***
	(0.017)	(0.016)
Worker Fixed Effects	No	Yes
Observations	45,135	45,132
Adjusted R-squared	0.006	0.258
Mean for Other Cycle Days	0.208	0.208

Notes: This table tests for sorting based on social networks. We use the daily worker panel to examine whether workers are more likely to work near their friends on cycle day 1, which was a workday where plot assignment was not randomized. Note that we measure friendships as reported during a baseline survey (before randomization). The estimates reported are based on a regression where the dependent variable is an indicator for working near a friend and the independent variable of interest is the coefficient on a dummy for cycle day 1. The omitted group in the regression is other cycle days, on which plot assignments were randomized.

Table A3: Additional Balance Tests: Comparing Worker Characteristics to Peer Ability

	Dependent Variable						
	Age	*Female*	*Married*	*Experience*	*Household Size*	*New Worker*	*HH Spending per Capita*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Log(Mean Peer Ability)	-0.257	0.091	-0.097	0.495	-0.044	-0.017	62.794
	(1.33)	(0.062)	(0.063)	(1.07)	(0.164)	(0.031)	(351)
Log(Leave-One-Out Gang Mean Ability)	-3.226	5.742***	-4.939***	-55.027***	-6.880***	0.865***	1,665.492
	(13.1)	(0.787)	(0.773)	(11.7)	(1.60)	(0.277)	(3521)
Cycle Day 1	No	No	No	No	No	No	No
Remaining Cycle Days	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Worker Fixed Effects	No	No	No	No	No	No	No
Date by Location Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	32,914	32,914	32,914	32,914	32,914	32,914	32,914
Adjusted R-squared	0.035	0.026	0.010	0.060	-0.006	0.067	0.003

Notes: This table presents results from regressions of a worker's characteristics on the mean ability of physically nearby co-workers. The underlying dataset is a panel at the worker and day level. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A4: Peer Effect Estimates Controlling for Other Peer Characteristics

	Dependent Variable: Log of Daily Output								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log(Mean Peer Ability)	0.030**	0.030**	0.037**	0.043***	0.040**	0.037**	0.034**	0.030**	0.040**
	(0.014)	(0.014)	(0.015)	(0.017)	(0.017)	(0.016)	(0.016)	(0.014)	(0.017)
Number of Neighbors		0.000							-0.001
		(0.001)							(0.001)
Mean Peer Age			0.000						0.000
			(0.000)						(0.000)
Share of Peers who are Female				0.009					0.016
				(0.007)					(0.013)
Share of Peers who are Married					-0.004				0.006
					(0.008)				(0.013)
Mean Peer Household Size						0.001			0.001
						(0.003)			(0.003)
Mean Peer Experience							0.001		0.001
							(0.000)		(0.000)
Share of Same-Gender Peers								-0.002	-0.001
								(0.007)	(0.007)
Worker Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date by Location Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	35,545	35,545	34,301	34,301	34,301	34,301	34,301	35,517	34,301
Adjusted R-squared	0.715	0.715	0.721	0.721	0.721	0.721	0.721	0.715	0.721

Notes: This table presents results from a regression of (log) daily output (kilograms of tea plucked) on the (log) mean ability of physically nearby co-workers, including controls for the other characteristics of workers' peers. The underlying dataset is a panel at the worker and day level. The measure of daily output comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by project staff. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A5: Effects of Direct and Strictly Second-Order Peers

	Dependent Variable: Log of Daily Output		
	(1)	(2)	(3)
Log(Mean Peer Ability)	0.030**	0.036***	0.037***
	(0.014)	(0.014)	(0.014)
Log(Mean Ability of Strictly 2nd-Order Peers)			0.000
			(0.019)
Worker Fixed Effects	Yes	Yes	Yes
Date by Location Fixed Effects	Yes	Yes	Yes
Observations	35,545	35,291	35,291
Adjusted R-squared	0.715	0.716	0.715

Notes: This table presents results from a regression of (log) daily output (kilograms of tea plucked) on the (log) mean ability of physically nearby co-workers. The underlying dataset is a panel at the worker and day level. The results in Columns (1) replicate our preferred specification from Table 3. The results in Column (2) are based on the restricted sample of observations where information on second-order neighbors is available. Strictly second-order neighbors are defined as the co-workers adjacent to a focal worker's neighbors who are *not* directly adjacent to the focal worker. Columns (3) and (4) estimate versions of Equation 1 that include measures of ability for strictly second-order neighbors. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field.

Table A6: Effects of Assigned Peers on Attendance and Tea Plucking

	Dependent Variable: Attendance	*Dependent Variable: Tea Plucking*
	(1)	(2)
Log(Mean Peer Ability)	0.003	0.004
	(0.012)	(0.013)
Worker Fixed Effects	Yes	Yes
Date-by-Location Fixed Effects	Yes	Yes
Observations	47,959	47,959
Adjusted R-squared	0.133	0.234

Notes: This table presents results from a regressions of an indicator for the worker being present at work (column 1) or being engaged in tea plucking (column 2) on the (log) mean ability of the physically nearby co-workers for their assigned field for the day. The underlying dataset is a panel at the worker and day level. The measure of daily attendance and plucking comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by project staff. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A7: Effects of Workplace Peers without Double Leave-One-Out Correction

	Dependent Variable: Log of Daily Output	
	(1)	(2)
Log(Mean Peer Ability)	0.053***	0.043***
	(0.014)	(0.014)
Worker Fixed Effects	Yes	Yes
Date Fixed Effects	Yes	No
Location (Field) Fixed Effects	Yes	No
Date by Location Fixed Effects	No	Yes
Observations	35,641	35,545
Adjusted R-squared	0.396	0.715

Notes: This table presents results from a regression of daily output (kilograms of tea plucked) on the mean ability of physically nearby co-workers. The underlying dataset is a panel at the worker and day level. The results in Columns (1) and (2) use two different approaches to control for date and location effects. The measure of daily output comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by staff for this project Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A8: Male and Female Peer Effects, Robustness

	Dependent Variable: Log of Daily Output	
	(1)	(2)
Male X Log(Mean Peer Ability)	0.006	0.020
	(0.020)	(0.041)
Female X Log(Mean Peer Ability)	0.060***	0.083*
	(0.023)	(0.046)
Quartiles of Age		
[Age 20 to 29] X [Log(Mean Peer Ability)]		0.031
		(0.045)
[Age 30 to 35] X [Log(Mean Peer Ability)]		-0.014
		(0.045)
[Age 36 to 44] X [Log(Mean Peer Ability)]		-0.015
		(0.042)
Quartiles of Own Ability		
[Own Ability Quartile 1] X [Log(Mean Peer Ability)]		-0.063
		(0.045)
[Own Ability Quartile 2] X [Log(Mean Peer Ability)]		-0.015
		(0.041)
[Own Ability Quartile 3] X [Log(Mean Peer Ability)]		0.004
		(0.041)
Worker Fixed Effects	Yes	Yes
Date by Location Fixed Effects	Yes	Yes
Observations	33,010	33,010
Adjusted R-squared	0.725	0.725
Male-Female Treatment Effect Difference	0.054*	0.063*
	(0.030)	(0.037)

Notes: This table presents results from a regression of (log) daily output (kilograms of tea plucked) on the (log) mean ability of physically nearby co-workers, interacted with worker characteristics. The specification includes a larger set of interaction terms relative to the model used for Table 4. This limits us to including full-saturated terms in just one of the interactions (gender). The underlying dataset is a panel at the worker and day level. The measure of daily output comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by project staff. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A9: Male and Female Peer Effects, Interaction Model Results

	Dependent Variable: Log of Daily Output	
	(1)	(2)
Log(Mean Ability of Female Peers)	0.024***	
	(0.008)	
Log(Mean Ability of Male Peers)	0.023***	
	(0.008)	
Male X Log(Mean Ability of Female Peers)		0.013
		(0.010)
Female X Log(Mean Ability of Female Peers)		0.037***
		(0.012)
Male X Log(Mean Ability of Male Peers)		0.012
		(0.011)
Female X Log(Mean Ability of Male Peers)		0.036***
		(0.012)
Worker Fixed Effects	Yes	Yes
Date by Location Fixed Effects	Yes	Yes
Observations	26,911	26,394
Adjusted R-squared	0.728	0.728

Notes: This table presents results from a (log) regression of daily output (kilograms of tea plucked) on the (log) mean ability of physically nearby co-workers, interacted with own and peer worker characteristics. The underlying dataset is a panel at the worker and day level. The measure of daily output comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by project staff. Sample is restricted to observations where gender is non-missing for all neighbors. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field.

Table A10: Additional Analysis of Peer Effects by Gender

	Dependent Variable: Log of Daily Output											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Log(Mean Peer Ability) X												
[Age 20 to 29]	0.025	0.092**										
	(0.037)	(0.045)										
[Age 30 to 35]	-0.029	0.081**										
	(0.039)	(0.040)										
[Age 36 to 44]	-0.009	0.042										
	(0.036)	(0.046)										
[Age 44 to 72]	0.043	0.024										
	(0.043)	(0.045)										
[Own Ability Quartile 1]			-0.052	0.033								
			(0.050)	(0.040)								
[Own Ability Quartile 2]			0.028	0.045								
			(0.040)	(0.040)								
[Own Ability Quartile 3]			0.003	0.115**								
			(0.039)	(0.047)								
[Own Ability Quartile 4]			0.024	0.061								
			(0.030)	(0.049)								
[Quartile 1 Exp.]					0.014	0.085*						
					(0.041)	(0.047)						
[Quartile 2 Exp.]					-0.018	0.065						
					(0.039)	(0.044)						
[Quartile 3 Exp.]					-0.007	0.056						
					(0.040)	(0.041)						
[Quartile 4 Exp.]					0.031	0.023						
					(0.038)	(0.043)						
[New Worker (=0)]							0.012	0.056**				
							(0.020)	(0.023)				
[New Worker (=1)]							-0.071	0.120				
							(0.077)	(0.114)				
Log(Non-Friends Mean Peer Ability)									0.008	0.050**		
									(0.019)	(0.022)		
No Non-friends (=1)									0.015	0.120		
									(0.082)	(0.105)		
Log(Friends Mean Peer Ability)									-0.005	0.026		
									(0.017)	(0.021)		
No Friends (=1)									-0.007	0.108		
									(0.074)	(0.084)		
Log(Mean Peer Ability)											0.005	0.062**
											(0.021)	(0.027)
Log(Mean Peer Ability), t-1											-0.015	0.018
											(0.023)	(0.025)
Log(Mean Peer Ability), t-2											0.014	-0.016
											(0.021)	(0.024)
Log(Mean Peer Ability), t-3											-0.006	0.000
											(0.021)	(0.024)
p -value: equal effects across gender	0.212		0.212		0.170		0.121		0.169		0.198	
Worker Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Date by Location Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,010		33,010		33,010		33,010		33,048		31,982	
Adjusted R-squared	0.725		0.725		0.725		0.725		0.725		0.726	

Notes: This table presents estimates from models that extend on the results from Tables 4–7. Each pair of columns reports results from a single regression where we interact the independent variables of interest from our prior tables with dummies for being male or female. Each column reports the point estimates for the gender-specific interaction terms. For example, Column (1) reports the coefficients from the interaction terms between a male dummy variable, log mean peer ability and a worker's own age group. Column (2) reports similar estimates for female workers. Sample is restricted to observations where worker gender is non-missing. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field.

Table A11: Summary Statistics for Worker Ability by Gender

	(1) Average	(2) Std. Deviation	(3) 10th Percentile	(4) 25th Percentile	(4) 50th Percentile	(4) 75th Percentile	(4) 90th Percentile	(5) Obs (N)
Overall	62.83	18.75	41.94	49.00	59.29	73.98	88.98	909
Females	58.07	16.87	40.25	45.76	54.26	67.21	81.93	393
Males	66.46	19.31	43.51	52.04	63.94	79.65	92.58	516

Notes: This table presents descriptive statistics for worker ability, for the subset of workers who have gender information from our survey data (909 of the overall total of 999 workers in our sample). The ability measure is estimated using Equation 2.

Table A12: Test for Asymmetry in Peer Effects

	Dependent Variable: Log of Daily Output	
	(1)	(2)
Log(Mean Peer Ability)	0.030**	
	(0.014)	
Log(|Peer Ability Difference|)X(Below Peer Ability)		0.017***
		(0.004)
Log(|Peer Ability Difference|)X(Above Peer Ability)		-0.018***
		(0.005)
Below Peer Ability		0.008
		(0.016)
Worker Fixed Effects	Yes	Yes
Date Fixed Effects	No	No
Location (Field) Fixed Effects	No	No
Date by Location Fixed Effects	Yes	Yes
Observations	35,545	35,449
Adjusted R-squared	0.715	0.717

Notes: This table presents results from a regression of (log) daily output (kilograms of tea plucked) on the (log) mean ability of physically nearby co-workers. The underlying dataset is a panel at the worker and day level. The results in Column (1) replicate our preferred specification from Table 3. Column (2) shows a test for asymmetry in the peer effects, following the analysis of the impact of friends in Bandiera, Barankay, and Rasul (2010). We compute the absolute value of the difference between the worker's own ability and the mean of their co-workers. We interact the log of this measure with an indicator for whether the worker's ability level is below the mean of their co-workers. The measure of daily output comes from administrative data obtained from Lujeri Tea Estates; information on neighbors was recorded and collected by project staff. Standard errors are two-way clustered at the level of a worker-by-cycle-day and date-by-field and adjusted using the Bayesian parametric bootstrap of Mas and Moretti (2009).

Table A13: Summary Statistics for Mean Ability of Peers who are Friends and Non-Friends

	(1) Average	(2) Std. Deviation	(3) 10th Percentile	(4) 25th Percentile	(4) 50th Percentile	(4) 75th Percentile	(4) 90th Percentile	(5) Obs (N)
Non-Friends	61.28	13.03	46.89	51.75	58.87	69.38	79.23	35,456
Friends	64.23	18.85	43.18	50.18	61.59	75.47	90.25	7,409

Notes: This table presents descriptive statistics for mean peer ability among workers' friends and non-friends. The ability measure is estimated using Equation 2.

Table A14: Preferences for Fast Peers by Gender

	(1)	(2)	(3)	(4)	(5)
	Males	Females	Raw Difference	Covariate-Adjusted Difference	Obs.
Who do you want to be reassigned next to?					
A fast plucker in your gang	0.70	0.75	0.057	0.115	434
A slow plucker in your gang	0.05	0.04	-0.008	-0.014	434
Any person of your choosing	0.09	0.09	0.003	-0.003	434
No reassignment	0.16	0.11	-0.052	-0.098**	434
Wants to work near a fast plucker and…					
...is willing to give up 1 bar of soap	0.59	0.59	-0.007	0.031	434
...is willing to give up 2 bars of soap	0.48	0.44	-0.046	0.056	434

Notes: This table presents statistics from survey data that we collected for a subset of tea pluckers at the Lujeri Tea Estates. Statistics are separated based on respondent gender. In the survey questions, fast and slow peers were described as coworkers who are in the top or bottom 10 percent of the gang in terms of kilograms of tea plucked per day, respectively. Column (3) shows the raw differences between the male and female rates for each outcome. Since gender is correlated with observed characteristics that affect worker preferences, Column (4) presents covariate-adjusted versions, computed by regressing the outcome on an indicator for being female, as well as controls for age, an indicator for being married, number of people in the worker's household, per-capita household expenditures, and years of experience. Appendix F provides details of the survey prompt and questions that we used to collect responses. For the choice experiment, respondents were given a gift of two bars of soap (worth 18 percent of average daily wages) and asked if they would be willing to give up soap in exchange for being reassigned.

B Bayesian Parametric Bootstrap

To account for the fact that neighbor (ability) types are estimated, we construct all standard errors in the paper using the Bayesian parametric bootstrap of Mas and Moretti (2009). This procedure consists of four steps. First, we draw simulated ability types (i.e., fixed effects for productivity) for each worker from a joint normal distribution that has a vector of means and a variance-covariance matrix that are equal to the results from our type estimation procedure. Second, for each draw of the worker types, we re-run the regressions in our analysis using the draws from the simulation to construct our peer ability measure. Third, we estimate σ_{sd} for a given regression as the standard deviation of the point estimates across draws. Fourth, we combine the across-simulation standard error with the typical clustered standard error. Let SE_{clust} be the usual cluster-adjusted standard error. The Bayesian parametric bootstrap standard error is then equal to $SE_{bayes} = \sqrt{\widehat{\sigma_{sd}}^2 + SE_{clust}^2}$. We follow Mas and Moretti in using 10 draws for our analysis.

C Double Leave-One-Out Approach

This section provides further details of our approach to addressing spatial spillovers, which may bias estimates of peer effects. Consider worker i who is on a high-productivity part of a field f, and let k index her K plot neighbors on that day. Suppose we estimate the following model of peer effects:

$$y_{ift} = \mu_i + \beta \overline{Ability}_{-if,t} + \delta_{tf} + \epsilon_{ift}. \tag{C1}$$

A potential problem occurs because the higher-productivity part of field f increases the error term ϵ_{ift}. We can control for field fixed effects in the model through the field-by-date fixed effects δ_{tf}. However, this does *not* control for variations in plot quality *within* a field.

In this case, the issue is that being on a high-quality part of a field f will increase the average value of output for worker i and all of worker k's neighbors (i.e., both y_{ift} and y_{kft} go up). This becomes a problem if we attempted to estimate ability for each worker using the specification:

$$y_{ift} = \mu_i + \mathbf{M}_{ift}\gamma' + \delta_{tf} + \tau_{ift} \tag{C2}$$

where the term $\mathbf{M}_{ift}$ is a vector of dummy variables which indicate whether worker j in a gang is working next to worker i in field f on date t. The issue in Equation C2 is that spatial spillovers will be absorbed in all of the neighbor k fixed effects (μ_k). This is problematic if one constructs the mean ability of the worker i's neighbors as $\overline{Ability}_{-if,t} = \overline{\widehat{\mu}}_{-ift} = \frac{1}{K}\sum_{k=1}^{K} \widehat{\mu}_k$. With this definition, spatial spillovers would generate a correlation between $\overline{Ability}_{-if,t}$ and the error term (ϵ_{ift}) in Equation C1.

Our approach eliminates this correlation by estimating each of the $\widehat{\mu}_k$ values for field f using a dataset that excludes field f. Specifically, we estimate Equation 2, reproduced below:

$$y_{igt} = \mu_{i-f} + \mathbf{M}_{igt}\gamma' + \delta_{tg} + \tau_{igt} \tag{2}$$

where g indexes all fields except field f. The term μ_{i-f} is the ability measure for worker i using all fields except field f. This allows us to construct a double leave-one-out estimate of a worker's mean peer ability, $\overline{Ability}_{-i-f,t} = \overline{\widehat{\mu}}_{-i-ft} = \frac{1}{K}\sum_{k=1}^{K}\widehat{\mu}_{k-f}$. This version of mean peer ability ($\overline{Ability}_{-i-f,t}$) is preferable because it is constructed without using any data from field f. Hence, spatially-correlated variation in plot quality within a field f does not affect our estimates of permanent productivity (ability type) for worker's peers.

One drawback of using this approach is that it will tend to induce additional classical measurement error in our ability estimates—and hence attenuate our estimated coefficients—because the ability measure is estimated using a smaller sample. In a situation with no spatially-correlated shocks, then, the approach that uses all the data to estimate ability is less biased.

To estimate the amount of additional measurement error induced by our approach, we first note that the average worker in our analytic sample has 5.61 cycle days of data. The double leave-one-out estimator drops one cycle day, and so uses 4.61 cycle days instead. This has the effect of using, on average, 1/5.61 fewer total observations to estimate the coefficients in the double leave-one-out approach. Since the standard error is proportional to $1/\sqrt{N}$, this inflates the standard error of the ability estimate by $\sqrt{5.61/4.61} = 1.103$. Thus, the variance of the ability estimate rises by 1.216. Under classical measurement error, the attenuation bias factor λ is $Var_x/(Var_x + Var_u)$, where x is the regressor of interest and the noise term u is additional variation in the measured value of x that arises as a result of using less data. The prior calculation shows that the double leave-one-out estimate has variance $Var_{x'} = 1.216 \cdot Var_x = Var_x + 0.216 \cdot Var_x$. In other words, the additional measurement error induced by the double leave-one-out approach implies that $Var_u = 0.216 \cdot Var_x$. Hence, the attenuation factor is $\lambda = Var_x/(Var_x + 0.216 \cdot Var_x) = 1/1.216$. This means that our estimates are attenuated by a factor of 0.822, and we should see results that are attenuated by an additional 18 percent.

The estimated attenuation factor is a lower bound, and thus the additional attenuation will be less than 18 percent, for two reasons. First, the ability estimates are correlated across days for a given worker, and so losing any given day is less costly in terms of mismeasurement of ability. Second, our preferred specification averages multiple independent worker ability estimates, which means that the calculations above overestimate the attenuation bias problem.[1] We study the attenuation bias issue numerically in Appendix D and find that our double leave-one-out estimates are attenuated by about 10 percent relative to the true parameter value.

[1] The above calculation also ignores the fact that the number of cycle days varies somewhat by worker, but this should not substantively affect our results.

D Monte Carlo Simulations

This section uses Monte Carlo simulations to assess the importance of two features of the empirical strategy described in Section 5. First, we show the sensitivity of the estimated impact of mean peer ability using different levels of fixed effects. The results show that our preferred estimates, based on including worker and field-by-date fixed effects, perform well relative to other choices of fixed effects. Second, we assess how the estimated impact of mean peer ability depends on the use of the double leave-one-out approach to estimating worker ability. For this exercise, we use simulations that assume the existence of spatially-correlated (time-invariant) shocks to plot quality. The results show that peer effect estimates based on the double leave-one-out estimator are much closer to the true parameter value relative to estimates based on an approach that estimates worker ability using all daily data.

D.1 Field-by-Date Fixed Effects

Our first set of simulations is based on the following data generating process for the log of daily output (measured in kilograms) y_{ift}:

$$y_{ift} = \mu_i + 0.1\overline{Ability}_{-if,t} + \delta_{tf} + \epsilon_{ift}. \qquad \text{(D1)}$$

We build the simulated data to mirror the following features of the observed sample of workers:

1. The simulated dataset has the same number of observations as our experimental dataset.

2. The date-by-field fixed effect δ_{tf} is estimated from the real data on productivity for a given field and date via our main regression specification (Equation 1).

3. Worker ability levels μ_i are drawn randomly, with replacement, from the estimated (log) ability levels for the entire gang that works on a given field.

4. Peers are determined by randomly assigning the simulated workers to positions in the real adjacency matrix for the field, which then determines log mean peer ability $\overline{Ability}_{-if,t}$.

5. The error term ϵ_{ift} is generated by randomly drawing a residual, with replacement, from the estimated distribution of residuals obtained from the main regression specification (i.e., Equation 1).

Under these conditions and the model in Equation D1, we create 150 simulated samples by re-generating all variables in each simulation iteration.

Using the simulated data, we use the double leave-one-out approach (as in the main text) to estimate the impact of mean peer ability using different sets of fixed effects. Appendix Table D1 (on page Appendix - 23) shows that estimates are much closer to the true parameter value when the peer effect regressions use worker and field-by-date fixed effects. The first row reports the true value of the peer effect parameter (which is constant across simulations). As expected, the second row shows that peer effect estimates are severely upward-biased when no fixed effects are included. The estimated coefficient is 0.55—five times the true value. This upward bias occurs because our experiment randomized workers to co-workers only within a given field. Fields vary systematically in terms of productivity, which will increase both daily output y_{ift} and estimated co-worker ability $\overline{\hat{Ability}}_{-if,t}$, leading to positive omitted variable bias. The third row reports the estimates from a model which only includes field-by-date fixed effects. There is evidence of exclusion bias due to the lack of worker fixed effects: workers cannot be their own co-workers, creating a mechanical negative bias in our point estimates (Guryan, Kroft, and Notowidigdo 2009; Caeyers and Fafchamps 2016). Finally, the fourth row shows that our preferred specification, which includes both worker and field-by-date fixed effects, produces estimates of the impact of mean peer ability that are close to the true parameter value. As discussed in Appendix C, the estimates are slightly attenuated due to classical measurement error.

Table D1: Monte Carlo Comparison of Peer Effects Estimates with Different Levels of FEs

	(1) Average	(2) Std. Deviation	(3) Min	(4) Max	(5) Obs (N)
True Peer Effect Coefficient	0.10	0.00	0.10	0.10	150
No Field-by-Date FEs, no Worker FEs	0.55	0.03	0.47	0.62	150
Field-by-Date FEs, no Worker FEs	0.01	0.04	-0.09	0.13	150
Preferred Estimates (Field-by-Date FEs and Worker FEs)	0.09	0.01	0.06	0.13	150

Notes: This table presents the distribution of the estimated coefficient on (log) mean peer ability across 150 Monte Carlo simulations of the data generating process described in Section D, with no spatially-correlated shocks. The first row shows the true parameter value within the simulation. Row 2 shows the estimates without any fixed effects in the model. Row 3 shows the estimates with field-by-date FEs but no worker FEs, and row 4 shows our preferred specification, which includes both field-by-date FEs and worker FEs. All specifications use the double leave-one-out estimator, which drops the cycle day in question from the sample when estimating the peer ability measure for a given day.

D.2 The Double Leave-one-out Approach

Our second set of simulations assumes the existence of spatially correlated shocks to plot quality. Specifically, this set of simulations is based on the following data generating process:

$$y_{ift} = \mu_i + 0.1\overline{Ability}_{-if,t} + \delta_{tf} + \sigma Shock_{if} + \sigma\rho PeersShocked_{if} + \epsilon_{ift}. \quad \text{(D2)}$$

The term $Shock_{if}$ is an indicator for worker i being shocked on field f; these shocks are permanent attributes of a specific plot, so each time a worker returns to a given field her shock status is the same. The term $PeersShocked_{if}$ is the number of peers near worker i that receive a shock. In this model, the idiosyncratic shocks to plot quality boost output by σ log points. The shocks can spill over onto directly neighboring plots as governed by the spatial correlation factor ρ; there are no second-order spillovers onto those plots' neighbors.

Based on this model, we create simulated data where we always assume that five percent of all plots receive a shock.[2] Across simulations, we vary two key parameters: the shock intensity, σ, which we allow to be 0, 0.25, or 0.5; and the spatial correlation factor, ρ, which we allow to be 0, 0.5, or 1. We simulate each combination of parameter values 50 times, randomly re-generating all variables for each simulation (holding the peer effect coefficient and shock rate constant).

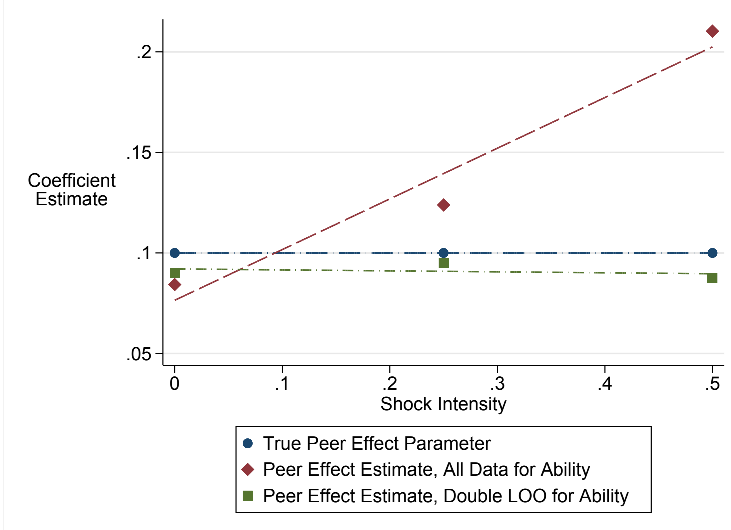
Using the simulated datasets, we compare estimates the impact of mean peer ability using the double leave-one-out approach to estimates based on an approach that uses all the data to estimate peer ability. All estimates of peer effects in this section are based on a regression which includes both worker and field-by-date fixed effects. Panel A of Appendix Figure D1 (on page Appendix - 26) shows estimates of the impact of mean peer ability while the shock intensity varies and the spatial correlation factor is held constant at 0.5. The bias of the estimator that uses all the data to estimate ability is an increasing function of the shock intensity. In contrast, the double leave-one-out estimator performs well even when

[2] We build the simulated data using all of the conditions listed in the prior section (e.g., the simulated sample has the same number of observations as the real data).

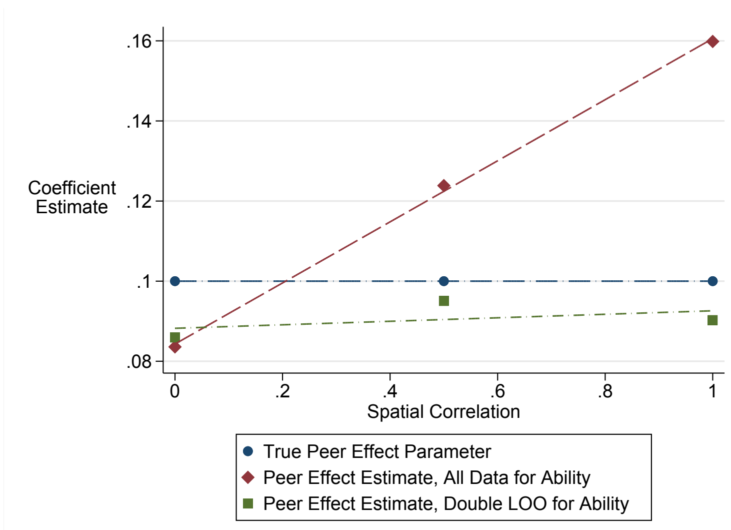
the shocks are larger. Next, Panel B of Appendix Figure D1 shows estimates that vary the spatial correlation factor and hold the shock intensity fixed at 0.25. Again, the double leave-one-out estimator has low bias for all values of the spatial correlation factor. In contrast, as the spatial correlation factor increases, the bias increases when using estimates of ability based on all the data.

Figure D1: Monte Carlo Simulation of the Performance of Different Peer Effects Estimators

(a) Varying Degree of Shock Intensity



(b) Varying Degree of Spatial Correlation



Notes: This figure presents estimates the impact of mean peer ability from different simulations of data. The "All Data" estimates (diamond symbol) are based on the simulation data and estimating worker ability using the entire sample. The "Double LOO" estimates (square symbol) are based on the simulation data and estimates of worker ability are based on a double leave-one-out approach where one cycle day is dropped. Panels A and B show points and regression lines where the underlying model for simulations varies by the shock intensity and spatial correlation parameters. See the text in Appendix D for further details.

E Details of Social Network Data Collection

As discussed in Section 4, we collected baseline social network data for workers in our analysis sample. For each worker, we asked for the first and last names of friends in two ways:

1. "Please tell us the names of good friends that are pluckers in your gang."

2. "Now tell us the names of people that you like chatting with at work that are pluckers that you have not already mentioned."

Enumerators electronically recorded all names provided (i.e., there was no maximum on the number of friendships provided).

F Details of Survey Questions on Preferences for Peers

As discussed in Section 4, we collected data on worker preferences on physically nearby peers. We collected this information using the following prompt and questions.

- **Introduction with general instructions:** I now would like to talk about your plot assignment on the different fields. As you noticed the capitao (supervisor) recently changed the plot assignments. What we want to do today is to ask everyone about their preferences over who to be assigned next to for the rest of this main season. For some respondents their preferences will be implemented starting next week. So, we will now ask you a series of 6 questions about who you want to work next to. In some of these questions we will ask if you would give up some of the bars of soap you will get as your respondent gift in order to work next to a person of your choice. And at the end of this interview the computer will give me a random number, similar to a lottery draw. The number determines whether you will be chosen for actual implementation. If you are chosen then we will pick one of your answers that you have given me and we will change your plot assignment according to the answer that you have given me. The chance of being chosen is 1 in 10. And if you are chosen we will pick ONLY ONE of your answers by chance for actual implementation. So you should answer the following questions as if you were chosen and consider one question at a time.

- **Reassignment question:** You have the opportunity to be reassigned next to a specific plucker you choose on all fields for the rest of the season. You can be reassigned to anyone you choose.

  Do you want to be assigned next to [respondents select one; the order of the options was randomized across respondents]:

  - You don't want to be reassigned
  - Name a specific person in the gang, for example a friend, or a relative, or anyone

else

- A slow plucker who is in the bottom [round(10%*gang size)] of your gang in terms of kg per day
- A fast plucker who is in the top [round(10%*gang size)] of your gang in terms of kg per day

- **Follow-up questions for respondents who chose "A fast pluckers. . . ":** Are you interested in being assigned next to a fast plucker who is in the top [round(10%*gang size)] of your gang in terms of kg per day—in exchange for ONE bar of soap from the respondent gift today?
  - Yes
  - No
- **If yes to the prior question:** Are you interested in being assigned next to a fast plucker who is in the top [round(10%*gang size)] of your gang in terms of kg per day—in exchange for TWO bars of soap from the respondent gift today?
  - Yes
  - No

Notes: Analogous follow-up questions are asked for respondents who chose "A slow plucker..." or "Name a specific person..." as responses. SurveyCTO instrument available on request.

References

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2010. “Social Incentives in the Workplace.” *The Review of Economic Studies*, 77(2): 417–458.

Caeyers, Bet and Marcel Fafchamps. 2016. “Exclusion Bias in the Estimation of Peer Effects.” NBER Working Paper 22565, National Bureau of Economic Research.

Guryan, Jonathan, Kory Kroft, and Matthew Notowidigdo. 2009. “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments.” *American Economic Journal: Applied Economics*, 1(4): 34–68.

Mas, Alexandre and Enrico Moretti. 2009. “Peers at Work.” *American Economic Review*, 99(1): 112–145.