TI 2014-051/I
Tinbergen Institute Discussion Paper

Peers at Work: From the Field to the Lab

Roel van Veldhuizen[1]
Hessel Oosterbeek[2]
Joep Sonnemans[2]

[1] *WZB Berlin Social Science Center, Berlin, Germany;*
[2] *Faculty of Economics and Business, University of Amsterdam, and Tinbergen Institute, The Netherlands.*

Electronic copy available at: http://ssrn.com/abstract=2430717

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance is a collaboration of the Dutch financial sector and universities, with the ambition to support innovative research and offer top quality academic education in core areas of finance.

DSF research papers can be downloaded at: http://www.dsf.nl/

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Electronic copy available at: http://ssrn.com/abstract=2430717

Peers at work: from the field to the lab*

Roel van Veldhuizen Hessel Oosterbeek Joep Sonnemans

Abstract

In an influential study, Mas and Moretti (2009) find that "worker effort is positively related to the productivity of workers who see him, but not workers who do not see him". They interpret this as evidence that social pressure can reduce free riding. In this paper we report an attempt to reproduce the findings of Mas and Moretti in a lab experiment. Lab experiments have the advantage that they can shut down alternative channels through which workers can influence the productivity of colleagues whom they observe. Although the subjects in our experiment are aware of the productivity of others and although there is sufficient scope for subjects to vary their productivity, we find no evidence of the type of peer effects reported by Mas and Moretti. This suggests that their findings are less generalizable than has been assumed.

Key words: peer effects, experiment, laboratory experiment.
JEL Classification: C91, J24

1 Introduction

Recent years have seen a large growth in the number of papers that try to address the extent to which important results can be generalized from lab experiments to the field (Camerer, 2011; Falk and Heckman, 2009; DellaVigna, 2009; Onderstal, Schram, and Soetevent, 2013). However, as Falk and Heckman (2009) note "the issue of generalizability of results is universal to all [empirical methods]" (page 537). Therefore, in this study we instead examine whether a specific very influential finding generalizes *from the field to the lab*.

For many field studies, interpreting data often requires several additional assumptions that exclude alternative mechansims. In contrast, laboratory experiments are able to control for such alternative mechanisms by design. More generally, lab experiments that exclude these alternative explanations provide

*This version: April 2014. Van Veldhuizen: WZB Berlin Social Science Center, Reichpietschufer 50, 10785 Berlin, Germany (email: roel.vanveldhuizen@wzb.eu, phone: +493025491402). Oosterbeek: University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Sonnemans: University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands and the Tinbergen Institute.

powerful support for the interpretation of existing field studies when they obtain similar results. However, when an experimental design that only allows for the essential elements (according to the interpretation of the field data) gives different results, we know that other aspects of the field setting must have played a role.

The field study we examine in this paper is Mas and Moretti (2009, abbreviated as "MM 2009" in the remainder of this paper). MM (2009) use a dataset on the productivity of cashiers (number of items checked per minute) of a large supermarket chain and show that cashiers in their supermarkets are influenced by the productivity of their coworkers. More specifically, the key result of their study is that "worker effort is positively related to the productivity of workers who see him, but not workers who do not see him" (MM (2009), p. 112). They interpret this finding by arguing that social pressure can internalize free riding externalities, but social preferences seem to be irrelevant. The study is considered to be very influential and its findings are often cited without reference to its specific setting (Ellingsen and Johannesson, 2008; Fehr, Goette, and Zehnder, 2009; Moretti, 2011; Charness and Kuhn, 2011)[1].

The extent to which MM (2009)'s setting reflects a larger class of production processes is ultimately an empirical question. Replications of MM (2009) in other field settings can inform us about this. In practice such replications are difficult to conduct because of the level of detail required in the data. Replicating the results of MM (2009) would require a data set that contains information about the relative spatial positioning of different workers (i.e., which workers can observe which coworkers) as well as a detailed individual-specific measure of production.

Yet even finding such datasets would not solve the problem that without experimentally controlling the situation one can never be sure that the interpretation of the data is the right one. For example, a cashier may put in more effort when observed by an experienced (productive) cashier who is likely to become a superior one day. In that case the relation between effort and the productivity of the observing worker is only indirect and we may interpret strategic behavior as a peer effect. More generally, in field settings the social context is richer and workers know more than only the productivity of the other workers. Other characteristics (unobserved by the researcher) that correlate with productivity may be the real driver of the effects. Yet knowing and understanding the exact mechanism behind the result is very important in particular when it comes to generalizability. Only in an experimental setup in which we can be sure that

[1]For example, Ellingsen and Johannesson (2008) state that MM (2009) shows that "low-productivity workers put in more effort when observed by high-productivity peers."

roles are randomized and the only information workers have is productivity we can really be sure that productivity is driving the effect.

Other aspects of MM (2009) can also be better controlled in the lab. For example, in an experiment the position (role) of the subject is determined exogenously, while in a supermarket an extra cashier may have some discretion to choose which counter to open. In the lab the ability (i.e., productivity) of the worker can be measured independently, in a non-social context. Communication is controlled in the lab and the experimenter can make sure that productivity is the only information workers receive. By contrast, if a cashier has a problem to solve (e.g. a bar code is not recognized by the scanner), she can ask for help from another cashier. Logically, instead of trying to get the attention of the cashier in front (who is sitting with her back to her), she is likely to ask the cashier behind, the one that observes her. This help is probably more useful if the observer is a more productive worker and it is also in line with low productivity workers being more affected by (positive) peer effects (which may be an alternative explanation of the main finding of MM (2009)).[2] The field data set as used by MM is a great opportunity and their findings have rightfully gotten a great deal of attention, but lab experiments are needed to establish the effects in a better-controlled environment.

In the design of our experiment we mimic the essential features of the work setting of the cashiers of the supermarket chain analyzed by MM (2009). Participants in the role of workers are members of teams and are not financially dependent on the effort of other workers but instead face a higher workload if their team members exert lower effort. Workers perform a real effort task (adding numbers), we vary worker observability and we obtain a direct measure of (permanent) productivity. Importantly, we can directly exclude mechanisms that MM do not discuss or have to assume are unimportant, such as cashiers asking help from each other, endogenous seating arrangements, strategic behavior relative to experienced coworkers, etc. Additionally, we vary the extent to which permanent productivity is visible to coworkers in the experiment, which allows us to differentiate between peer effects based on contemporareous production and baseline productivity, something that MM (2009) are unable to do.

In line with the results of MM (2009), we expected workers to work harder when partnered with more productive coworkers who can observe them during

[2]MM (2009) discuss this possibility on page 134-5. They find that productive workers are not negatively affected by observing unproductive coworkers and argue that this would have been the case if they were helping their improductive colleague. However, the costs in time of providing a bar code are neglegible, whereas the benefits for the improductive coworker are large.

the production process. However, we find that this result does not carry over to the laboratory context. We observe no evidence of peer effects despite the finding that workers are well aware of the production and productivities of their coworkers when this information is available and have the ability to both increase and decrease their effort levels. This suggests that MM (2009)'s findings may not be as general as has been assumed.

The remainder of this paper is structured as follows. The next section gives a brief overview of the results of MM (2009) and discuss other related studies. Section 3 describes the experimental design and Section 4 formulates our hypotheses. Section 5 presents and discusses the results and the final section concludes.

2 Overview of Mas and Moretti (2009)

MM (2009) estimate peer effects among a two year sample of 394 cashiers working for a US-based supermarket chain. As a measure of productivity, they use the average number of items scanned by a cashier over a 10 minute interval. This abstracts away from several potentially important aspects of performance (such as quality of service and absenteeism), though it does provide a precise estimate of cashiers' production speeds. For every 10 minute interval and every supermarket, the authors know exactly which workers were on duty and at which cash register they were working. This allows them to identify the workers' spatial orientation, which they use to define the observing and observable sets.

At this point, it is useful to first give a few definitions of terms used by MM (2009) that we will also use in this study. The *focal worker* is the worker in a shift whose behavior is being analyzed. The *observing set* consists of coworkers who are facing (and can thus see) the focal worker. The *observable set* consists of coworkers whom the focal worker is facing.

The empirical estimates are based on variations on the following equation (equation 6 in M&M, 2009).

$$\Delta y_{itcs} = \alpha + \beta \Delta \overline{\theta}_{-itcs} + \pi \Delta N_{tcs} + e_{itcs} \tag{1}$$

Here, Δy_{itcs} is the production of worker i in 10 minute interval t, at date c, at store s and relative to worker i's production in the previous 10 minute interval. $\Delta \overline{\theta}_{-itcs}$ is the change in the average permanent productivity of coworkers, ΔN_{tcs} is the change in the number of workers on duty and e_{itcs} is the error term. Peer effects are captured by the coefficient β; a positive coefficient indicates that workers increase their production speed when the average permanent pro-

ductivity of their coworkers increases. The key independent variable -average permanent productivity of coworkers- is not directly observed and needs to be constructed using a separate estimation procedure that corrects for possible influences of arbitrary social interactions.

Note that equation (1) assumes that peer effects operate through permanent productivity. However, in practice workers may also be influenced by the contemporaneous effort of their coworkers. MM (2009) are not able to empirically distinguish between these two mechanisms, since any attempt to estimate a model that includes contemporaneous effort as an independent variable would have led to a reflection problem (Manski, 1993).[3] MM (2009) argue that their estimates are likely a combination of both the contemporaneous and permanent productivity effects.

MM (2009)'s main findings are presented in Table 1. The first column shows the results for equation (1). The coefficient indicates that increasing the average permanent productivity of the coworkers by 10% increases the production of focal worker i by 1.5%. Column 2 shows that these peer effects only appear for changes in the average permanent productivity of coworkers who can observe the focal worker (the observing set). This suggests that peer pressure is more important than prosocial preferences. Column 3 shows that peer effects are specific to workers with below average permanent productivity.

MM (2009) report several additional estimates. These show inter alia that peer effects persist over time and are larger for workers who occupy registers in the close proximity of the focal worker. Most relevant for this study are the results reprinted in column 4 of Table 1, which show that adding a worker to the observing set increases the focal worker's productivity by 3.1%. Interestingly, adding a worker to the observable set decreases the focal worker's productivity by almost the same percentage.

We would like to mention here two other recent papers about peer effects in the lab which are related to MM (2009) but have a different setup than our paper. Beugnot et al. (2013) use a multiplication task (multiplying two-digit numbers by one-digit numbers) and subjects are paid according to a piece-rate. In the baseline treatment subjects are observed (by subjects in later sessions) but apparently without their knowledge. Subjects in the recursive treatment observe the performance of the subjects of the baseline treatment, but are not observed themselves. In the simultaneous treatment players observe and are observed by other players in the same session. In all cases, observation does not happen continuously but rather at the end of every experimental period.

[3]The reflection problem is the identification problem that arises when trying to identify a causal relationship between two individuals (or groups) that are observed simultaneously.

Table 1: Main Results of Mas and Moretti (2009)

Dependent Variable: Difference in log productivity of the focal worker between t and $t-1$				
	(1)	(2)	(3)	(4)
Δ Average coworker permanent productivity	.15			
	(.02)			
Δ Average coworker permanent productivity in observing set		.17		
		(.02)		
Δ Average coworker permanent productivity in observable set		.01		
		(.02)		
Δ Average coworker permanent productivity (below average worker)			.24	
			(.03)	
Δ Average coworker permanent productivity (above average worker)			.05	
			(.04)	
Δ Presence worker in observing set				.031
				(.003)
Δ Presence worker in observable set				-.030
				(.003)
Number of cashiers	394	394	394	394
Observations	1,718,052	1,649,916	1,718,052	1,732,941

Notes. This table displays the results of four regressions taken from Mas and Moretti (2009). Columns (1), (2), (3) and (4) are column (1) of table 2, column (1) of table 6, column (1) of table 3 and column (3) of table 6 in Mas and Moretti (2009) respectively. For column (3), the coefficient for above average workers is the total coefficient for above average workers, which is the sum of the two coefficients reported by Mas and Moretti (2009) in column 1 of table 3. The observing set consists of workers who are facing the focal worker. The observable set consists of workers whom the focal worker is facing. The bracketed numbers are standard errors.

This design is certainly interesting but misses the category we are particularly interested in: players who know that someone of either high or low productivity observes them.

Georganas, Tonin, and Vlassopoulo (2013) use a slider task (the objective is to position as many sliders as possible at exactly 50) and pay a piece-rate on an individual or team basis. Teams have three members: one control who does not observe and is not observed, one observer and one observed player. The team members did not play simultaneously; first the observed player (and the control) worked on the task while the observers were seeing on their screen a copy of the computer screen of their observed team member. Only after the observed players finished, the observer performed the task himself. This design differs from ours in five important ways: (1) their design does not include a player who observes and is also observed; (2) the productivity of observers is unknown to the observed player; (3) subjects are paid a piece-rate; (4) observing players are not simultaneously working on the task and finally (5) the way subjects are observed is very different. Observers directly observe the actions of the observed player, which allows them to learn how to better approach the task from the observed player. All peer effects Georganas, Tonin, and Vlassopoulo (2013) find appear in the first few periods, which indicates that their study can tell us more about peer effects on *learning* than on *effort* (which of course can be interesting in its own right).

3 Experimental design

The goal in designing the experiment was to capture the essential features of the work setting of the supermarket chain analyzed by MM (2009) while maintaining a high level of experimental control. The experiment consisted of two stages. In the first (or baseline) stage, participants worked alone, allowing us to obtain a measure of their permanent (or baseline) productivity in the absence of peer effects. In the second (or production) stage, participants worked in teams of four, allowing us to investigate the impact of peer effects. In the remainder of this section we first discuss the baseline stage, followed by a discussion of the production stage and a brief comment on the procedure at the end of the experiment.

3.1 Baseline stage

The experiment was computerized using PhP/MySQL. In the experiment, participants had to perform a production task that consisted of adding three two-

digit numbers. We chose this task since it is easy to understand, captures the essential features of the production process described by MM (2009) and results in sizable differences in productivity between participants, which allows us to examine differences between low productivity and high productivity workers. The three numbers appeared on the computer screen together with information about whether the answer to the previous exercise was correct and the cumulative number of successfully completed exercises up to that point. The sequence of numbers used in the exercises was randomly generated before the first session of the experiment, so that it was identical for all participants; we used a separate sequence for the baseline stage and the production stage.

Upon entering the laboratory, participants were welcomed and assigned to a random computer. They received the instructions for the baseline stage of the experiment on screen; the instructions included a single check-up question. After everyone had finished the instructions, the baseline stage started. In the baseline stage, participants worked individually for 4 minutes and were paid 10 Euro cents for every correct answer they provided.[4]

3.2 Production stage

After the baseline stage, the experiment moved to the production stage, for which participants received additional instructions and check-up questions. After all participants had finished the instructions and check-up questions, the production stage started. In the production stage, there were three different treatments. In two treatments (BASELINETEAM and TEAM), participants were randomly grouped into teams of four. They were told that as a team they had to solve a number of exercises somewhere between 750 and 1150 (the actual number was 829). There was no longer a fixed production time; participants received a fixed fee of 10 euros for their participation in the production stage, regardless of the number of exercises they had solved. They were told that they would be paid and could leave after the team had finished. Thus, similar to MM (2009) workers were not financially dependent on the effort of their coworkers, but face a higher workload if their coworkers are less productive.

Importantly, they were also told that during the production stage they might receive information about the number of exercises solved by one or more of their teammates. The left part of their computer screen contained an overview of their team. Figure 1 gives the team overview used for treatment BASELINETEAM. An arrow going from one participant to another indicates that

[4]An English translation of all instructions and two screenshots can be found in the appendix; the original Dutch version of the instructions is available upon request.

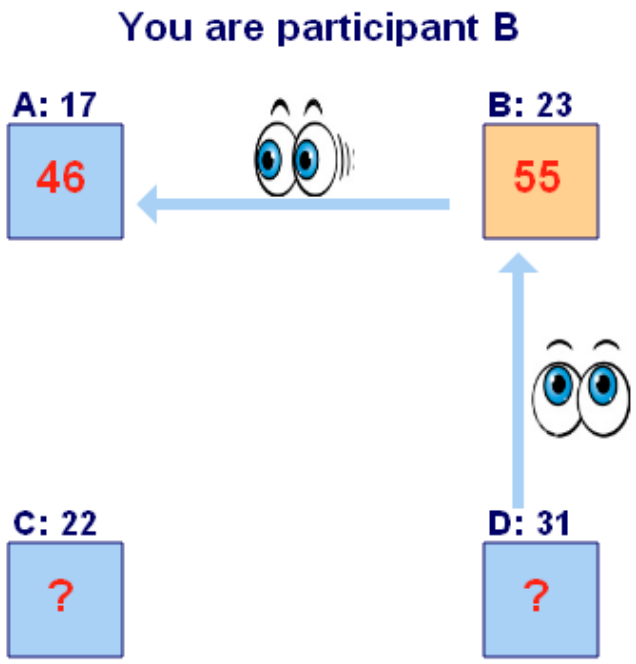


Figure 1: Team Overview

Notes. The figure gives the team overview used in treatment BASELINETEAM of the experiment. The numbers above the squares are baseline productivities. These are visible for all team members in treatment BASELINETEAM and are absent in treatment TEAM. The numbers inside the squares are the cumulative production levels in the production stage; each participant only knows her own production and the production of participants she can see (as indicated by the arrows).

this participant could see the number of exercises solved by the other in the production stage. For example, in our set-up participant B could see the number of exercises solved by participant A.

The team structure we used allows us to compare four different information perspectives. Participant A knew the number of exercises solved by him could be seen by one team member, whereas participant D knew he could see the number of exercises solved by one team member. Participant B knew he could both see one team member and be seen by another team member and participant C knew he could neither see nor be seen by another participant. The structure of the team remained fixed for the duration of the experiment.

The difference between treatments BASELINETEAM and TEAM is that participants in treatment BASELINETEAM also learned the number of exercises solved by their team members in the *baseline* part of the experiment (their permanent productivity). They learned the permanent productivity for all participants in their team, even for those for whom they did not know the

number of exercises solved in the production stage. As a consequence, treatment BASELINETEAM is closest to MM (2009) in that it allows peer effects to work through permanent productivity as well as contemporaneous productivity; treatment TEAM only allows peer effects to work through contemporaneous productivity. Treatment TEAM allows us to provide a stronger test of MM by allowing us to measure if peer effects work even when workers do not know the productivity of their coworkers directly. Note that there is no reflection problem in either treatment, since information flows go only in one direction: for example worker B can be influenced by the contemporaneous productivity of worker A but not vice versa.

Finally, we also ran an individual treatment in which participants individually had to solve between 188 and 288 exercises (the actual number was 207). Participants in treatment INDIVIDUAL never got any feedback about the performance of other participants in the experiment and were also allowed to leave the experiment after having solved the required number of exercises. We included this treatment to check if organizing workers into teams per se changed their productivity.

3.3 End of the experiment

For each participant, the production stage ended after she (treatment INDIVIDUAL) or she and her team (treatments BASELINETEAM and TEAM) had completed the required number of exercises. After finishing their final exercise, participants received an overview of their earnings and were asked to fill out a questionnaire. The questionnaire contained several demographic questions, a self-monitoring questionnaire (Snyder, 1974) and questions about the experiment. After finishing the questionnaire, participants could collect their payment and leave the laboratory, even if other participants were still solving exercises or working on the questionnaire.

In total we ran 9 sessions in February and April 2012, in which 188 subjects participated (84 in TEAM, 84 in BASELINETEAM and 20 in INDIVIDUAL). Participants had an average age of 22.5, 38% reported they studied economics and 58% were male. Subjects earned between 17.60 and 22.50 euro, with an average of 19.45 euro. The average number of exercises solved per minute did not differ between the individual treatment (6.85) and the two group treatments (6.61), so we will focus on the two group treatments in the analysis.

4 Hypotheses

MM (2009) find that increasing the average permanent productivity of the set of coworkers increases workers' production speeds. Importantly, they find that this effect only appears when the change applies to coworkers who are in the focal worker's observing set and that this effect only appears among low-productivity workers. For the experiment, these findings directly translate into three hypotheses:

- Hypothesis 1 (peer effects): Increasing the average permanent productivity of the set of coworkers increases the production speed of the focal worker.
- Hypothesis 2 (observability): Increasing the average permanent productivity of the set of *observing coworkers* increases the production speed of the focal worker; increasing the average permanent productivity of the set of *observable coworkers* has no effect.
- Hypothesis 3 (ability): Increasing the average permanent productivity of the set of coworkers increases the production speed of a low productivity focal worker, but not of a high productivity focal worker.

All three hypotheses are testable using treatment BASELINETEAM. Hypothesis 1 could in principle also apply to treatment TEAM. Although permanent productivity is not directly available to workers in treatment TEAM, workers B and D do know the contemporaneous production of one team member; they could infer the permanent productivity of this team member from his contemporaneous production speed. Hypothesis 2 can only be tested using treatment BASELINETEAM, since permanent productivity levels of observing coworkers are not known and cannot be indirectly inferred in treatment TEAM.

5 Results

We can only expect to find peer effects if subjects in the experiment are aware of the information they are supposed to observe and if the task in the experiment allows for sufficient variation in workers' production speeds. The first subsection presents evidence indicating that these two requirements are satisfied. We then turn to the main results of the paper.

Table 2: Recall of production stage

	Overall production of player			
	A	B	C	D
Player A's estimate	.93	.12	-.26	.07
Player B's estimate	.89	.85	.21	.17
Player C's estimate	-.11	-.26	.93	.27
Player D's estimate	-.12	.70	.44	.91

Notes. This table displays Pearson correlation coefficients between the actual production of a player and the estimates given by all four players. The numbers on the diagonal represent the workers' recall of their own production. 6 workers did not fill out the recall questions and were thus omitted from the sample, leaving 162 observations.

Table 3: Recall of permanent productivity

	Permanent productivity of player			
	A	B	C	D
Player A's estimate	.94	.82	.48	.71
Player B's estimate	.96	.98	.94	.83
Player C's estimate	.85	.87	1	.82
Player D's estimate	.70	.78	.94	.95

Notes. This table displays Pearson correlation coefficients between the permanent productivity of a player and the estimates given by all four players. The numbers on the diagonal represent the workers' recall of their own permanent productivity. 4 workers did not fill out the recall questions and were thus omitted from the sample, leaving 80 observations.

5.1 Awareness and variation

Subjects can only respond to their coworkers' productivity if they are aware of it. To assess whether this requirement is fulfilled, we asked subjects to recall the overall production of their coworkers and (in treatment BASELINETEAM) their permanent productivities. These questions were not incentivized and were not announced until after the experiment had ended. Table 2 shows that worker recall of the overall production of observable coworkers was very good (the correlation is between .7 and .9 and significant at the 1% level in all cases). Moreover, Table 3 shows that workers in BASELINETEAM also did well at recalling the permanent productivity of their coworkers; all correlations are positive, large and significant at the 1% level. This suggests that workers were well aware of both the permanent productivity and (when applicable) the current production of their coworkers.

To respond to the productivity of coworkers, subjects need to be able to adjust their production speed. We therefore examine whether workers were able to change their production speed across exercises. We do not look directly at minutes spent per exercise, since some exercises were more difficult than

Table 4: Variation in production speed

Standard deviation of production speed	.451	.495	.425	.372
Weighted standard deviation of production speed	.412	.436	.402	.350
Sample	All	BASELINETEAM	TEAM	INDIVIDUAL
Observations	188	84	84	20

Notes. This table gives the average standard deviation of worker production speed relative to the average production speed per exercise. For the second row, the standard deviation is weighted by the production speed of the respective worker relative to the average production speed in the experiment.

others. Instead, we look at worker production speed per exercise divided by the average production speed for the respective exercise (among all workers). Table 4 shows that the average standard deviation in worker production speed is around 40-50%, which suggests that workers had sufficient scope to adjust their production speed.[5]

5.2 Peer effects

Hypothesis 1 states that workers' production speed is increasing in the average permanent productivity of their coworkers. As a measure of production speed, we take the average number of exercises solved by the worker per minute in the production stage. As a measure of permanent productivity, we take the number of exercises solved by the worker in the baseline stage.[6] For the dependent variable, we divide production speed by permanent productivity to correct for differences in permanent productivity.[7]

Table 5 shows the result of an OLS regression of the log of worker production speed (relative to the baseline) on average coworker permanent productivity. The results in the first column are based on both treatments. The point estimate equals -0.059 with a standard error of 0.084, implying that we do not find a significant effect of the permanent productivity of coworkers on the focal

[5]More anecdotally, we observed on several occasions that workers simply stopped working for a few minutes, suggesting that decreasing the production speed was quite possible as well.

[6]As an alternative measure of permanent productivity, we also considered taking only a subset of the baseline, for example only the last two minutes. We chose the overall baseline production since it was most highly correlated (r=.78) to production speed in the production stage. In any case, using subsets of the baseline leads to very similar results.

[7]In alternative specifications, we also corrected for permanent productivity by including it as an additional covariate. The results of this approach are identical to the approach presented here. We also ran the regressions in levels, which yielded qualitatively identical results as well.

Table 5: Peer effects estimates

Dependent Variable: Log average production speed per minute of the focal worker relative to baseline			
	(1)	(2)	(3)
Log average coworker permanent productivity	-0.059	-0.017	-0.090
	(0.084)	(0.146)	(0.090)
Constant	0.206	0.126	0.271
	(0.148)	(0.254)	(0.162)
Sample	all	BASELINETEAM	TEAM
Observations	126	84	42

Notes. This table displays the results of three OLS regressions; the numbers in parentheses are standard errors. The dependent variable is the log of the average number of exercises solved per minute in the production stage divided by the average number of exercises solved in the baseline stage per minute. Average coworker permanent productivity is the average permanent productivity of all coworkers for which the permanent productivity can be observed. For treatment BASELINETEAM, this is the average permanent productivity of all three coworkers. In treatment TEAM, coworker permanent productivity could not be observed directly. However, workers B and D were indirectly aware of the permanent productivity of workers A and B respectively through seeing their production speed in the production stage. Thus, for treatment TEAM we use the permanent productivity of workers A and B for focal worker B and D respectively; workers A and C are not included in the analysis.

worker productivity. The results in columns (2) and (3) for the separate treatments are very similar. Recall that MM find a larger, strongly significant and positive effect of average coworker productivity. By contrast, we find a *small*, *insignificant* and *negative* peer effect in this study.

5.3 Observability

Thus we find no evidence that peer effects are relevant at the aggregate level. However, it is possible that this obscures the fact that peer effects are active more locally. Indeed, by hypothesis 2 peer effects should be larger with respect to coworkers who can observe the focal worker. Table 6 displays the results of three regressions that examine if this is indeed the case. The regressions examine if the focal worker's production speed is affected by the permanent productivity of the observing coworker or the observable coworker. In both cases, we use only the results for BASELINETEAM, since permanent productivity is only available to workers in this treatment.

The results (column 1) show that increasing the permanent productivity of the observing coworker by 10% *decreases* the production speed of the focal worker by 0.61%. This effect is considerably smaller in absolute size than

Table 6: Peer effects estimates by observability

Dependent Variable: Log average production speed per minute of the focal worker relative to baseline	(1)	(2)
Log permanent productivity (observing set)	-0.061	
	(0.147)	
Log permanent productivity (observable set)		-0.094
		(0.145)
Constant	.238	0.232
	(0.252)	(0.242)
Sample	Players A&B	Players B&D
Observations	42	42

Notes. This table displays the results of three OLS regressions. The regressions use data from treatment BASELINETEAM only. The numbers in parentheses are standard errors. The dependent variable is the log of the average number of exercises solved per minute in the production stage divided by the average number of exercises solved in the baseline stage per minute. Average coworker permanent productivity is the average permanent productivity of all three coworkers.

the effect found by MM (2009) (1.7%), goes in the opposite direction and is not significant. Increasing the permanent productivity of *observable* coworkers decreases the production speed of the focal worker by 0.94%. This effect is considerably *larger* than MM (2009)'s effect and of the opposite sign, though similar to MM (2009) it is not significant. Moreover, the estimated coefficients of observable and observing coworkers are very similar, suggesting that there is no difference between being observed by or observing a coworker, again in contrast to MM (2009)'s results.

5.4 Ability

Thus far we have found no evidence of peer effects at the aggregate level or separately for observing or observable coworkers. By hypothesis 3, one reason for the lack of effect could be that peer effects only appear among low productivity workers. To investigate if this is the case, we repeat the regression of table 5 separately for low and high productivity workers (using a median split on baseline productivity).

Table 7 shows the results of these regressions. Increasing the average permanent productivity of coworkers by 10% increases the production speed of high productivity workers by 2.09%, whereas it reduces the production speed

Table 7: Peer effects estimates by ability

Dependent Variable: Log average production speed per minute of the focal worker relative to baseline	(1)	(2)
Log average coworker permanent productivity	-0.081	0.209
	(0.185)	(0.183)
Constant	0.345	-0.382
	(0.322)	(0.317)
Productivity	low	high
Sample	BASELINETEAM	BASELINETEAM
Observations	44	40

Notes. This table displays the results of four OLS regressions; the numbers in parentheses are standard errors. The dependent variable is the log of the average number of exercises solved per minute in the production stage divided by the average number of exercises solved in the baseline stage per minute. Average coworker permanent productivity is the average permanent productivity of all three coworkers.

of low productivity workers by 0.81%. Thus, if anything high productivity workers appear more likely to be positively affected by peer effects, although neither coefficient is significant at conventional levels. By contrast, MM (2009) found that low productivity workers were significantly affected by peer effects, whereas high productivity workers were not affected by peer effects at all.[8]

6 Discussion

This study reports the results of a laboratory experiment to test the generalizability of the results found in an influential field study (Mas and Moretti (2009)). The other way around is much more common: a field study that tries to test the external validity of laboratory experiments (see e.g., Gneezy and List, 2006; Onderstal, Schram, and Soetevent, 2013; Armantier and Boly, 2008 or see DellaVigna, 2009; or Camerer, 2011 for a recent literature review). A field study is less intrusive and more natural than a laboratory experiment, but pays the price of missing the extent of control that is possible in a laboratory experiment (Camerer, 2011).

[8]We also ran four other sets of analyses which are not directly related to our hypothesis. These showed that (1) production speed is similar for all player types, (2) production speed does not differ between the three treatments, (3) the estimates for contemporaneous peer effects are very similar to the estimates reported in this section, (4) there is no evidence that peer effects differ for low and high self-monitoring types (Snyder, 1974). More detailed results are available upon request.

The experiment discussed in this study should not be considered a replication of Mas and Moretti (2009) in the strict sense of the word: for that comparable field data would be needed. Instead, we designed a laboratory experiment that contains all the elements that are supposed to be the driving forces behind their results. In particular, we used a repetitive real effort task where we varied whether workers were observing or being observed and the information subjects had about the productivity of their coworkers was carefully controlled.

Overall, the results of Mas and Moretti (2009) do not generalize to the experiment discussed in this study. In particular, we find no evidence of peer effects and also see no evidence that workers are more likely to be influenced by the productivity of coworkers in their observing set. We have also shown that this result is not due to individual heterogeneity in ability and appears despite the finding that workers were well aware of the characteristics of their coworkers and were able to substantially vary their production speed.

One might argue that we fail to find peer effects in our experiment due to a lack of precision. The estimate in the first column of Table 5, for example, only excludes peer effects exceeding 0.11 with 95% probability. This argument ignores, however, that the information that is supposed to trigger social pressure is much more salient in our laboratory setting than it can be in any field setting in which the productivity of others may be unknown and in which there are many other factors (e.g., customers) that may distract cashiers from the alleged peer pressure

Our results thus suggest that other aspects of Mas and Moretti (2009)'s specific field setting (that they could not control and that are absent in our lab experiment) may have caused their effects. For example, a cashier who encounters a problem can ask help from another cashier and will ask this help typically from a cashier in whose line of view she is. In a laboratory experiment we have more control and such effects can be ruled out.

On a more general level, our study suggests that there may be an important role for laboratory experiments to study how general the results of influential empirical field papers are. To the extent that empirical results reflect fundamental aspects of human behavior, they should be observed in a wide variety of contexts, including the laboratory.

Acknowledgments

We gratefully acknowledge comments from seminar participants at the University of Amsterdam, at the ESA 2012 in New York and the fall 2012 BBE workshop in Berlin. We thank Speerpunt Behavioral Economics at the University of Amsterdam for financial support.

References

Armantier, Olivier and Amadou Boly. 2008. "Can corruption be studied in the lab? Comparing a field and a lab experiment." *CIRANO Working Papers* .

Beugnot, Julie, Bernard Fortin, Guy Lacroix, and Marie Claire Villeval. 2013. "Social Networks and Peer Effects at Work." IZA discussion paper No 7521.

Camerer, Colin F. 2011. "The Promise and Success of Lab-Field Generalizability in Experimental Economics: A Critical Reply to Levitt and List." *Working Paper* .

Charness, Gary and Peter Kuhn. 2011. "Lab labor: what can labor economists learn from the lab?" In *Handbook of Labor Economics Volume 4A*, edited by David Card and Orley Ashenfelter. New York: Elsevier, 229–330.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2):315–372.

Ellingsen, Tore and Magnus Johannesson. 2008. "Pride and Prejudice: The Human Side of Incentive Theory." *American Economic Review* 98 (3):990–1008.

Falk, Armin and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science* 326 (5952):535–538.

Fehr, Ernst, Lorenz Goette, and Christian Zehnder. 2009. "A Behavioral Account of the Labor Market: The Role of Fairness Concerns." *Annual Review of Economics* 1 (1):355–384.

Georganas, Sotiris, Mirco Tonin, and Michael Vlassopoulo. 2013. "Peer Pressure and Productivity: The Role of Observing and Being Observed." IZA discussion paper No. 7523.

Gneezy, Uri and John A List. 2006. "Putting behavioral economics to work: resting for gift exchange in labor markets using field experiments." *Econometrica* 74 (5):1365–1384.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies* 60:532–542.

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

Moretti, Enrico. 2011. "Local Labor Markets." In *Handbook of Labor Economics Volume 4B*, edited by David Card and Orley Ashenfelter. New York: Elsevier, 1237–313.

Onderstal, Sander, Arthur J.H.C. Schram, and Adriaan R. Soetevent. 2013. "Bidding to give in the field." *Journal of Public Economics* 105:72–85.

Snyder, Mark. 1974. "Self-monitoring of expressive behavior." *Journal of Personality and Social Psychology* 30 (4):526–537.

A Instructions

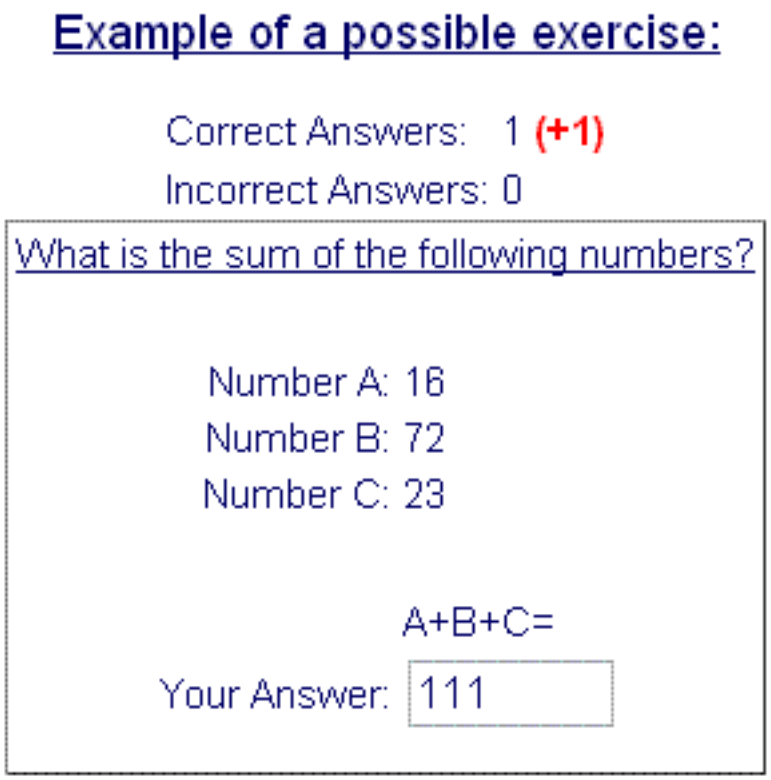
Welcome to this experiment. During the experiment you are not allowed to communicate with other participants. If you have a question, please raise your hand. One of the experimenters will then come to your cubicle to answer your question.

Today's experiment consists of two parts; part two will take considerably more time than part one. Part two of the experiment will be explained after you have finished part one of the experiment. Your income will be determined on the basis of your results in the experiment. You will also receive a show-up fee of 7 Euros.

Please read through the following instructions carefully. As part of the instructions you will be asked a practice question to test your understanding of the instructions. When you have correctly answered this question, the experiment will move on. Using the navigation bar at the top of your screen it will be possible to return to previous pages during the instructions and practice question.

Instructions Part One

In part one of the experiment the procedure will be as follows. The computer screen will display three two-digit numbers (as in the example below). Your task is to calculate the sum of these three numbers. For every correct answer you will receive 10 Euro cents. An incorrect answer does not earn you any money; after an incorrect answer you will automatically go on to the next exercise. This part of the experiment will take up 4 minutes in total; during these 4 minutes you can do as many exercises as you want. The clock in the lower right corner of the screen tells you how much time you have left. The number of exercises you have answered both correctly and incorrectly is displayed above the current exercise; the (+1) indicates if the previous exercise was answered correctly or incorrectly.



Practice Question

Hank has finished 11 exercises, providing the correct answer to 8 and an incorrect answer to 3. How many Euro cents has Hank earned?

Instructions End

You are now ready for part 1 of the experiment. By pressing the link below you will reach a waiting screen. The first part of the experiment starts as soon as all the others have also finished the instructions. On the waiting screen you can read back the text of the instructions.

Instructions Part Two

Like in part one, your task in part two will be to add three two-digit numbers. However, during this part of the experiment you will form a team with three other persons. The experiment will last until you and the other three people in your team have provided a correct answer to a fixed number of exercises. This fixed number of exercises will be somewhere between 750 and 1150 exercises. For this part of the experiment both you and all other team members will get a fixed payment of 10 Euros.[9]

[9]These are the instructions for treatment BASELINETEAM. The instructions for the other treatments are available on request.

As soon as you and your team have solved the required number of exercises, the experiment will be over for your team after a short questionnaire. One of the experimenters will come to your cubicle to pay out your earnings for the experiment. After payment you can leave the laboratory, even if the other teams are not done yet.

Information

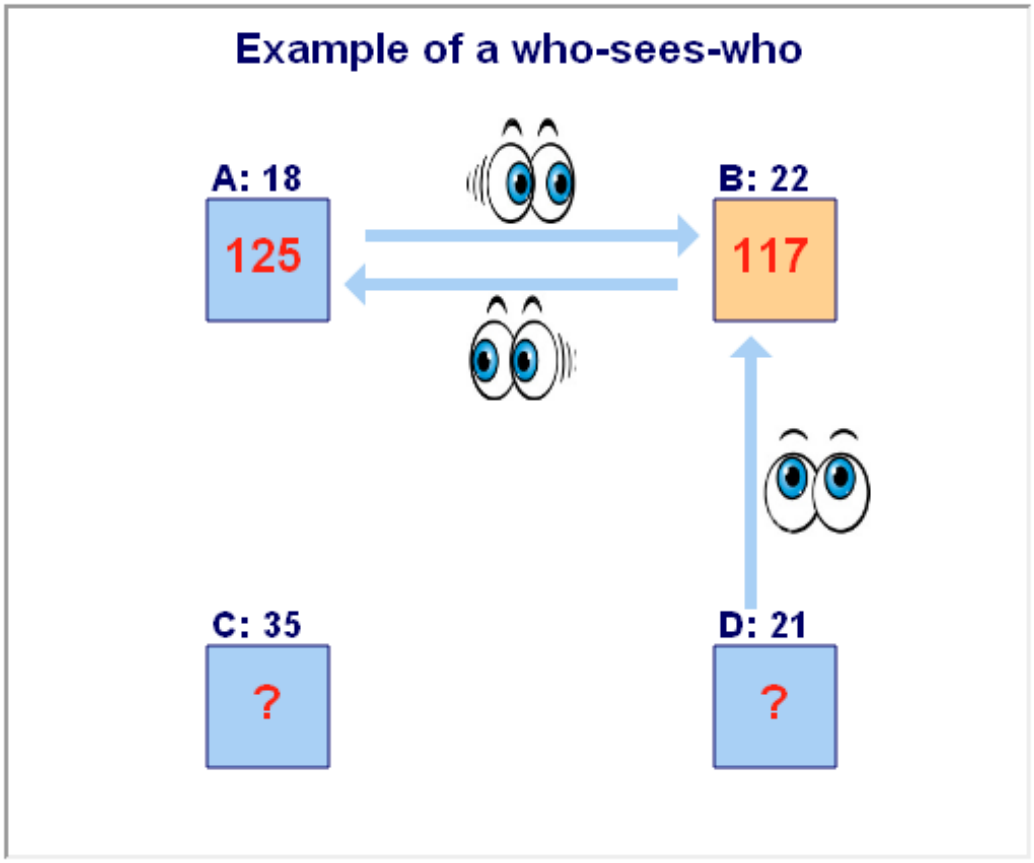
During the experiment the left side of your screen provides an overview of your team, comparable to the figure displayed below. Each of the squares A, B, C and D represents one of the team members; your square will be colored in orange (in the example below you are participant B). Within the figure, the blue numbers above the squares indicate how many exercises each participant has solved in part 1 of the experiment. Thus, in the example below, participant A has solved 18 exercises, participant B 22, participant C 35 and participant D 21.

The figure also contains arrows between some participants. In the example below there is an arrow from participant D to participant B, an arrow from participant B to participant A and an arrow from participant A to participant B. These arrows represent information flows. An arrow from one participant to another indicates that this participant is able to see the number of solved exercises of the other participant in part two up to that point. Only the number of correct answers will be counted. The number of solved exercises by other participants will be displayed in red letters within the square that corresponds to said participant.

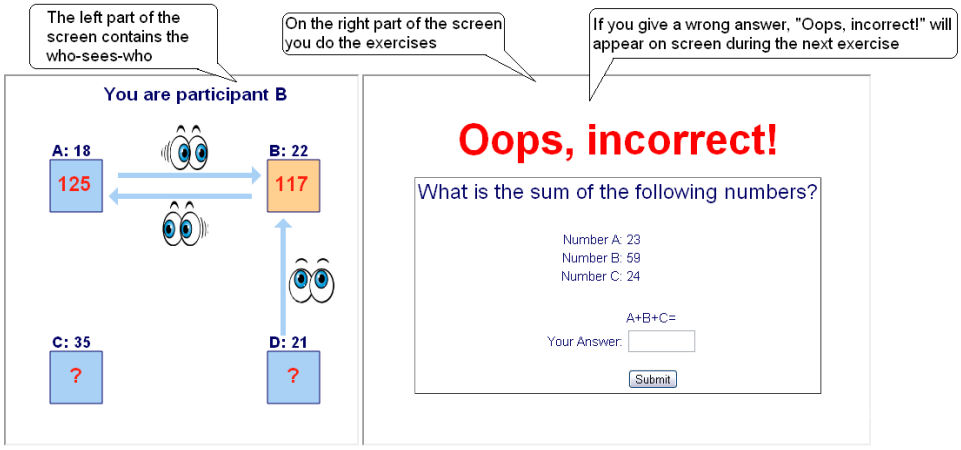
In the example below you are participant B and can see how many exercises you have solved (117) and how many exercises participant A has solved (125). In the example below you have **no** information about the number of exercises solved by participants C and D, who therefore have a "?" in their corresponding square. This means that you will at no stage get to know the number of exercises solved by participants C and D (not even after the experiment).

Finally, note that both participant A and participant D can see how many exercises you have solved up to that point. Participant C, however, does not know how many exercises you have solved and will at no stage get to know this number (not even after the experiment).

The figure below only represents an example of a who-sees-who; the who-sees-who that will be used in the experiment (which can have fewer, more or different arrows) will be announced after the instructions. However, the who-sees-who will remain the same during the experiment; id est, both the arrows, your participant letter and the participant letter of the other participants will remain the same for the whole of the experiment. The composition of your team will not change during the experiment either.


Example of a who-sees-who
A: 18
125
B: 22
117
C: 35
?
D: 21
?

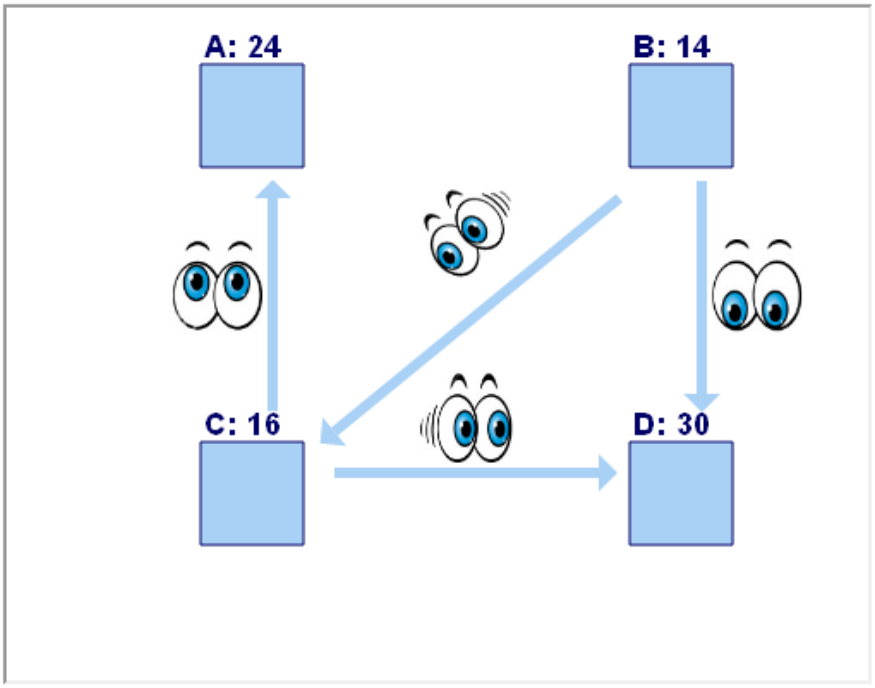
Example Screen



Check-Up Question 1

The figure below gives an example of a who-sees-who. Indicate for every team member for which team member he or she can see the number of solved exercises. Also indicate for every team member who can see the number of exercises solved by them.

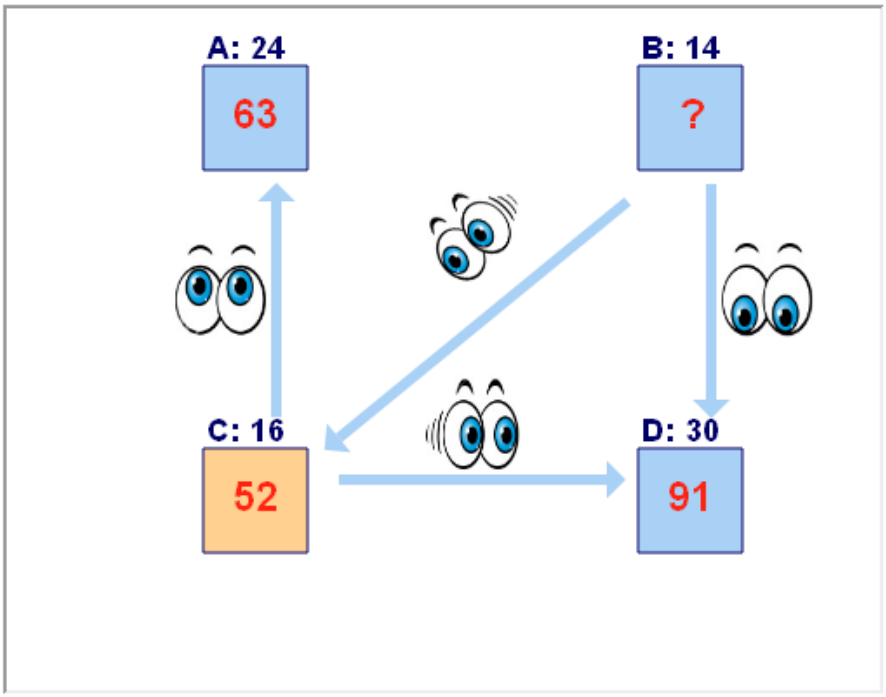
	Participant A	Participant B	Participant C	Participant D
Participant **A** knows the number of exercises solved by:	☐	☐	☐	☐
Participant **B** knows the number of exercises solved by:	☐	☐	☐	☐
Participant **C** knows the number of exercises solved by:	☐	☐	☐	☐
Participant **D** knows the number of exercises solved by:	☐	☐	☐	☐
This participant knows the number of exercises solved by participant **A**:	☐	☐	☐	☐
This participant knows the number of exercises solved by participant **B**:	☐	☐	☐	☐
This participant knows the number of exercises solved by participant **C**:	☐	☐	☐	☐
This participant knows the number of exercises solved by participant **D**:	☐	☐	☐	☐



Check-Up Question 2

The figure below gives an example of a who-sees-who (the same one as in the previous question). You are participant C, therefore you have all the information that participant C has access to. Indicate for all participants how many

exercises they have solved in part one of the experiment. When possible, indicate for every participant how many exercises this participant has solved so far in this part of the experiment (part two). If the number of solved exercises is not known, do not fill in anything.



Check-Up Question 3

Finish the sentence: this part of the experiment ends when *you/your team mates/you and your team mates/everybody in the experiment* have given the correct answer to *750/1150/a fixed number between 750 and 1150* exercises.

Check-Up Question 4

Which participants will know after the experiment how many exercises you have answered correctly?

- All participants in the experiment
- All participants who during the experiment could see the number of exercises you solved.
- All participants of which during the experiment you could see the number of exercises solved.
- Nobody

Instructions End

You are now ready to start part two of the experiment. By pressing the link below you will arrive at a waiting screen. Part two of the experiment will start as soon as all participants have finished the instructions. On the waiting screen you can read back the instructions of this part of the experiment as well.

B Screenshots

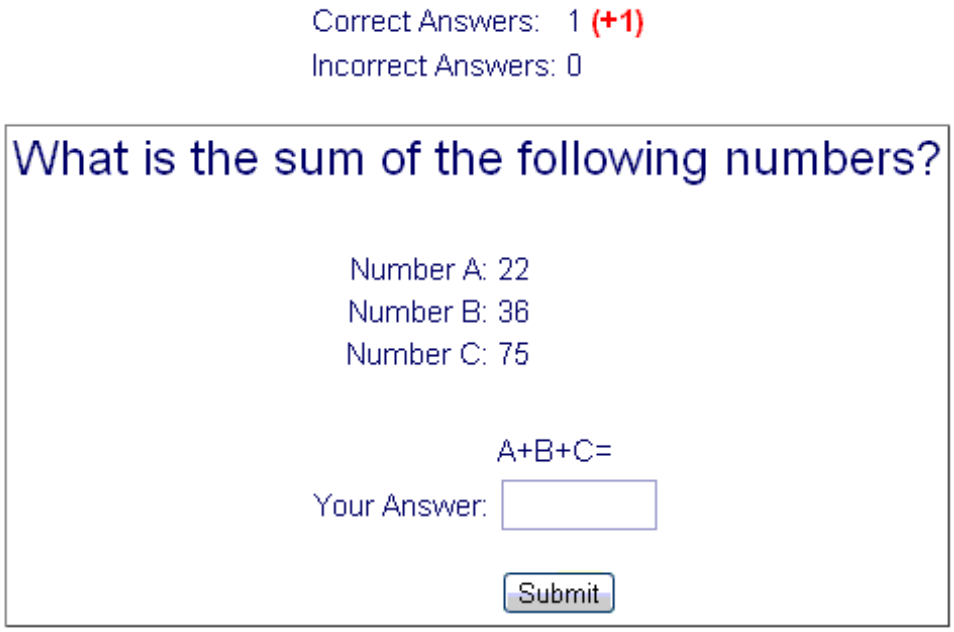


Figure 2: Screenshot of the Baseline Phase

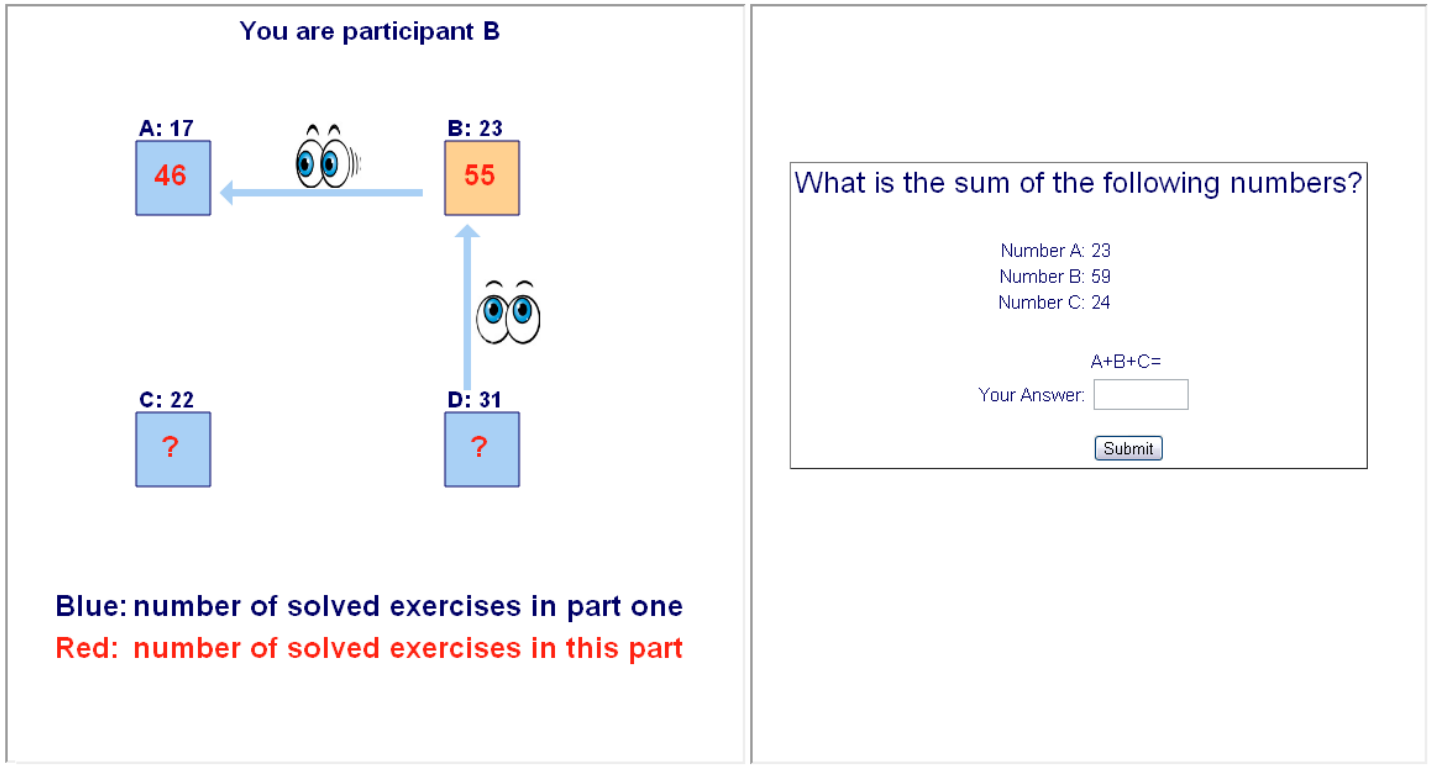


Figure 3: Screenshot of the Production Phase