The impact of formal incentives on teams: Micro-evidence from retail

Francisco Brahm, Antoine Feylessoufi, Marcos Singer[1]

Abstract

The impact of formal incentives on team productivity is not fully understood. A handful of field experiments have documented a small or null impact of incentives on teams, but could not rule out the possibility of incentives being ineffective on individuals to begin with. In this paper we estimate the impact of incentives both on individuals and on teams, within the same setting. We find that incentives improve productivity by 19% on individuals; however, they have a null effect on teams. We show that this null effect is not because incentives do not work—on the contrary, they strongly boost team productivity via effort complementarity—but because there is no need for them: teams without incentives display larger social incentives than teams with incentives.

Keywords: effort complementarity, free-riding, monitoring, pay for performance, productivity, social incentives, teams.

1 Introduction

Motivating workers to exert effort under moral hazard is a long-standing research area in the economics of organizations (Prendergast, 1999; Gibbons and Roberts, 2012; Lazear and Oyer, 2012). The use of formal incentives—both pay-for-performance (e.g., Lazear, 2000) and income-relevant monitoring systems (e.g., Pierce et al., 2015)—is a central tool to achieve this goal. Organizational design, as well as strategic management more broadly, is increasingly addressing formal incentives, studying how they affect important organizational and strategic phenomena such as change and inertia (Kaplan and Henderson, 2005), network structure (Mitsuhashi and Nakamura, 2022), the balance between exploration and exploitation (Lee and Meyer-Doyle, 2017), creativity and innovation (Ederer and Manso, 2013),

[1] Brahm: Assistant Professor, London Business School, 26 Sussex Pl, London NW1 4SA, UK, fbrahm@london.edu (corresponding author); Feylessoufi: Research Fellow, School of Management, University College London, One Canada Square, London E14 5AA, UK and Judge Business School, University of Cambridge, a.feylessoufi@ucl.ac.uk; Singer: Full Professor, School of Management, Pontificia Universidad Catolica de Chile, Av. Libertador Bdo. O'Higgins 440, 4to piso, Santiago 8331010, Santiago, Chile, msinger@uc.cl. We thank comments received by participants on earlier—and much different—versions of this paper at: brown bag seminar of the School of Management, Pontificia Universidad Catolica de Chile (2013); SMS Copenhagen special conference (2014); SMS Annual Conference in Berlin (2016); and the Madrid Work and Organization Workshop (2017). For the current version, we are grateful for the comments received from participants in the Sumantra Ghoshal Conference (2021) and seminar participants at the Strategic Organization Design (SOD) unit at the Danish Institute for Advanced Studies and the London Business School.

integration of effort among firm divisions/units (Kretschmer and Puranam, 2008) and collaboration (Lee and Puranam, 2017).

This paper studies how the use of formal incentives affects effort provision in teams; that is, we focus on the case when production is carried out by a group of workers, wherein individual contributions to aggregate output are not observable, and thus, not directly contractible (Alchian and Demsetz, 1972). Therefore, formal incentives translate into individual group members' compensation being a function of group output (Holmström, 1982).

Addressing this question is important for the following two reasons. First, the use of teams to carry out operations and execute important strategic initiatives has increased over time (Kozlowski and Bell, 2013; Garton and Noble, 2017). Therefore, improving our understanding of how formal incentives affect the performance of teams is of consequence. Second, although we have developed a rather sophisticated understanding of the impact of incentives on individual workers—the comparison B – A in Table 1—we still have much to learn in the case of teams—the comparison D – C in Table 1 (Bloom and Van Reenen, 2011; Lazear, 2018; Nyberg et al., 2018). Regarding the impact of pay-for-performance on individual workers, extensive evidence shows that it increases the effort of individual workers executing simple tasks that require physical effort (as opposed to creativity) and for which output is well defined and measurable (as opposed to multidimensional tasks) (for reviews, see Prendergast, 1999; Lazear, 2018; Lazear and Oyer, 2012); however, it may not work—and may even be detrimental—when tasks require intrinsic motivation (Frey and Oberholzer-Gee, 1997; Gubler et al., 2016), social comparisons are salient (Obloj and Zenger, 2017; Larkin et al., 2012), and workers can adapt to—and game—the incentive system, particularly over time (Frank and Obloj, 2014; Obloj and Zenger, 2019; Brahm and Poblete, 2018). Regarding monitoring, there is also evidence for its effectiveness in raising the effort of individual workers, particularly in settings where monitoring technology is imperfect and workers can therefore behave as rational shirkers (Nagin et al., 2002; Nalbantian and Schotter, 1997; Schweitzer et al., 2018; Jensen et al., 2020; Duflo et al., 2012).

In contrast, the impact of formal incentives on teams remains unclear (Bloom and Van Reenen, 2011; Prendergast, 1999; Lazear and Oyer, 2012; Lazear, 2018). Regarding pay-for-performance, Friebel et al. (2017) indicate: "unlike for the case of individual incentives, the jury on the effectiveness of team incentives is still out" (p. 2169). To the best of our knowledge, there are only three studies that cleanly estimate the impact of pay-for-performance in teams (versus a control of fixed wages[2]): Friebel et al. (2017) find 3% increase in productivity using a field experiment within a German retail chain of bakeries; Delfgaauw et al. (2020 and 2022) conduct field experiments in a retail chain of women lingerie/swimwear stores as well as a retail chain of outdoor clothing and shoes, finding no effect in both cases. As for monitoring in teams, we found no clean field evidence (cf. Grosse et al., 2011).

Table 1. Framework to assess the impact of formal incentives on productivity

		Formal contract leads to…	
		Weak incentives (fixed wage, low monitoring)	Strong incentives (pay-for-performance, high monitoring)
Production is carried out by…	Individual	*Productivity A*	*Productivity B*
	Team	*Productivity C*	*Productivity D*

This paper contributes to the literature in three ways. First, we provide a new estimate of D – C that addresses an issue that has not been considered in extant research. Prior work estimated D – C without also analyzing B – A in the same setting. However, there is no guarantee that the small or null effects estimated by Friebel et al. (2017) and by Delfgaauw et al. (2020 and 2022) are not due to a feature of their settings, such as organizational culture, that also renders formal incentives to individuals ineffective. Our unique setting allows us to estimate all four productivities: A, B, C and D (as shown in Table 1). Therefore, *we can make a comparison of the impact of formal incentives on individuals and on teams,*

[2] Several studies instead compare the performance of teams with incentives against the performance of individuals without incentives (Hamilton et al., 2003; Boning et al., 2007; Burgess et al., 2010; Hansen, 1997; Weiss, 1987; Gaynor and Gertler, 1995; Brosig-Koch et al., 2017; Hennig-Schmidt et al., 2011; Erev et al., 1993; Babcock et al., 2015; Bandiera et al., 2013; and Nalbantian and Schotter, 1997). These studies provide evidence of "the impact of adopting team production, conditional on strong incentives" (i.e., D – B in Table 3), and are sometimes cited—unfortunately—as providing evidence of "the impact of formal incentives on teams" (e.g., Bloom and Van Reenen, 2011; Nyberg et al., 2018).

holding constant the environment. To our knowledge, we are the first study to provide a full comparison of D – C vs. B – A in a single setting.

The second contribution is to decompose our estimate of D – C into the main generic mechanisms that past literature has highlighted as drivers of the impact of incentives on team effort: incentive-driven effort provision, free-riding, effort complementarity, and social incentives. Previous estimates of D – C have not provided a breakdown of the mechanisms involved. These mechanisms operate as follows. First, formal incentives motivate *incentive-driven effort provision*. Second, when compensation is tied to group outcome, individual team members can *free-ride* on other team-members' efforts, thus reducing their incentive to contribute effort and decreasing the effectiveness of the group pay-for-performance incentive (Homlstrom, 1982; Alchian and Demsetz, 1972). Third, there can be *effort complementarities* between team members; that is, effort by worker *i* improves the productivity of worker *j*, and thus of the whole team (Itoh, 1991; Puranam, 2018)[3]. The presence of team incentives induces each team member to (selfishly) boost their own effort in order to improve team output—and in turn, their own individual compensation. This effect can strongly counteract the free-riding problem. Fourth, work executed in teams is particularly geared to generate *social incentives,* which are defined by Ashraf and Bandiera (2018) as "any factor that (a) affects the marginal benefit or marginal cost of effort and (b) stems from interactions with others. Social incentives can be underpinned by either nonstandard preferences (altruism, reciprocity, etc.) or social interactions between selfish agents" (p. 440). For example, one prominent type of social incentive is "peer pressure" which occurs when a team member—usually one exerting high levels of effort—pressures their lagging teammates to increase their efforts (Kandel and Lazear, 1992; Mas and Moretti, 2009; Chan et al., 2014). Other types of social incentives include social comparison, status-seeking, altruism, reciprocity, relational contracting, group identity, and group

[3] Although less frequent, there may be effort substitutability, which occurs when individual contributions to the team are separable and the principal introduces relative pay-for-performance schemes (i.e., agent *i* gains more if they exert more effort, but also if other agents do worse).

cohesion. Social incentives tend to significantly increase effort; however, there are circumstances in which they may be detrimental, such as social comparisons that lead to envy and sabotage (Ashraf and Bandiera, 2018; Larkin et al., 2012). Crucially, social incentives usually (but not necessarily) become weaker (stronger) when formal incentives are present (absent) (Ashraf and Bandiera, 2018; Bowles and Polania-Reyes, 2012). For example, the meta-analysis by Herbst and Mas (2015) shows that peer pressure is crowded out (in) by pay-for-performance (fixed wage).

Our third contribution is to show that without having an estimate of B – A, *it is difficult to evaluate the magnitude or even the presence of the mechanisms* driving any estimate of D – C. This may explain why previous literature has lagged in unpacking the mechanisms described above. Intuitively, B – A facilitates the assessment of free-riding in teams: the extent of free-riding is measured, by definition, by how much the impact of pay-for-performance on effort decreases as team size increases, *in comparison to the effort level provided by the individual choosing effort alone, without the possibility to free-ride*. Thus, without B – A, it is not possible to have this baseline comparison, and thereby assess how much free-riding is being mitigated by the counteracting forces of effort complementarity and social incentives. This means that the mechanisms explored by Friebel et al. (2017) and Delfgaauw et al. (2020 and 2022) are suggestive and partial, because their research design cannot fully reveal what is going on "under the hood."

To guide our analysis, we set up a simple formal model, based on Itoh (1991), Kandel and Lazear (1992) and Ashraf and Bandiera (2018). Although it has been tailored to our setting, the model is general and flexible: it accommodates individual and team production, under either weak formal incentives (fixed/base and low monitoring) or strong formal incentives (pay for performance and high monitoring). The model incorporates the three generic mechanisms that the literature has identified as playing a role in the impact of team incentives on productivity: free-riding, effort complementarity, and social incentives. Social incentives are allowed to vary depending on the strength of formal incentives. Using the model's first-order condition of optimal worker effort, we can predict the productivities A, B, C and D of Table 1 and, therefore, the differences (B – A), (D – C), (C – A), and (D – B); moreover, we can assess the role of free-riding, effort complementarities and social incentives in each of these differences. We then take these predictions to the data.

We study the productivity of shelf replenishment in supermarkets for two beverage (e.g., soft drinks, water, beer) producers that, combined, account for more than 90% of the market. We have monthly firm-store level productivity data for roughly 175 stores for three years: 2011, 2014 and 2015. In small stores, each firm appoints one worker to execute replenishment; therefore, there is "individual production." In larger stores, firms appoint two or more workers (as many as 12 to 14 in the largest stores) who work two non-overlapping shifts: the morning shift and the afternoon shift. Store-level output is not separable among workers or shifts; thus, when there is more than one worker in the store, we are observing "team production" (Alchian and Demsetz, 1972; Holmström, 1982). Regarding contracting, one firm offers pay-for-performance and strong monitoring throughout all years. The other firm offers a fixed wage and weak monitoring in 2011 before pivoting to mimic its rival for the years 2014 and 2015 (the contractual change was designed and piloted, respectively, in years 2012 and 2013). This allows us to conduct a difference-in-differences estimation controlling for store-firm fixed effects.

For B – A, we find that weakly incentivized individuals (i.e., at the 10th percentile of store size, which have one worker) display 19% lower productivity than strongly incentivized individuals. In contrast, for D – C, we find that weakly incentivized teams at the 90th percentile of store size (which have four or more workers) display the same level of productivity as strongly incentivized teams. Given that incentives work in motivating individuals, one might be tempted to conclude that the null D – C is a product of free-riding. However, we can measure the size of effort complementarity by exploiting the case where there are two workers in the store who have non-overlapping shifts. This case, while displaying effort complementarity, strongly reduces the likelihood of social interaction and, thus, of social incentives. Guided by our model, we find that effort complementarity is large in our setting and, thus, that it strongly counteracts free-riding. Therefore, D – C is null not because D is declining (on the contrary, D is remains high due to complementarity). Instead, we show that D – C is null because C increases due to social incentives: guided by the model, we use the estimate of (B – A) – (D – C) to infer that the impact of social incentives on effort is approximately 17 percentage points greater under weak incentives than under strong incentives.

Overall, these results show that formal incentives do not increase team productivity in our setting—not because they do not work, but because teams do not need them. Formal incentives do motivate effort, but they crowd out the social incentives that emerge in non-incentivized teams.

For identification, we rely on many elements. First, we add period and firm-store fixed effects so that results are identified using variance within the store-firm pair. Second, the firms use the same contract across all stores, and thus, the match between incentive strength and the choice of individual vs. production technology—which is defined by store size—is exogenous. Third, regarding worker sorting into store size (and thus to individual vs. team production technology), we use a subsample to show that several worker observables do not vary across store size. Fourth, regarding sorting into firm characteristics (including incentive strength), we first indicate that workers are supplied by undifferentiated staffing/employment agencies such as Adecco and Randstadt and, therefore, that the main sorting choice of workers is whether to work for agencies; then, we show that the results do not change when the agency is shared between firms; this contradicts the notion of sorting, because sorting into firms is much easier within the same agency rather than across. Fifth, in order to control for other changes that might have been correlated with the contractual change, we show that our results are robust to the addition of several time-variant covariates; also, the test of bounds on omitted variable bias surpasses the recommended threshold (Oster, 2019). Finally, although we do not directly observe social incentives (our estimate of social incentives is a "residual" based on our model of workers' effort), we can nonetheless trace its presence by showing that, consistent with previous research, group pro-sociality under weak incentives is likely to be driven by increased threat from a rival group (Bauer et al., 2016; Francois et al., 2018; Bohm and Rockenbach, 2012); in our case, this threat would be an increase in the productivity of the rival replenishment team. For this analysis, we use an instrumental variable technique exploiting a headquarters relocation by one of the firms.

The rest of the paper is organized as follows. In section 2, we introduce the details of our setting. In section 3, we present the model and introduce the predictions we test. In section 4, we present our data and the empirical analysis. In section 5, we discuss and conclude.

2 Setting

2.1 Overview of replenishment operations in supermarkets

The supermarket shelf replenishment operations that we study are located in Santiago, the capital of Chile. Replenishment of store shelves is a key activity: it consumes approximately 40% of total supermarket costs, and shelf stock-outs hurt sales as well as the images of supermarkets and suppliers alike (see Corsten and Gruen, 2003, for a review; and Reiner et al., 2013, for a detailed process study). In Chile, replenishment is a labor-intensive activity executed at an hourly wage of approximately USD 3.50 (as of 2013). Large consumer product manufacturers hire replenishment workers to be at the stores. Most manufacturers do not hire directly; instead, they outsource to human resource agencies that are the workers' direct employers (e.g., Randstad, Adecco, Manpower). These agencies specialize in staffing and human resources services (e.g., payroll), offering an undifferentiated service.

Replenishment of beverages consists of four tasks: product ordering and product reception (which requires 15% of the allotted time), warehouse management (30%), shelf stocking (50%), and monitoring of commercial agreements and promotions (5%).

Replenishment presents a canonical agency problem: supervision at stores is difficult and imperfect, thus leading to low observability of worker effort. Supervisors, who can be hired by manufacturers and/or agencies, must regularly visit each store to evaluate work performance. However, a typical store is visited only once or (at most) twice per week, and remote monitoring technologies are not available.[4] Therefore, workers have considerable latitude in their effort decisions.

In small stores, there is typically one worker. In medium-sized stores, there are two or three workers working in two non-overlapping shifts of one or two workers per shift. In large stores, there are four or more workers, also in two non-overlapping shifts of two or more workers per shift. Output is non-separable between workers within a shift—it is unfeasible to apportion sales volume to specific workers or shifts. Thus, team production is in place in large stores and in medium stores, whereas individual production is in place in small stores.

There are two sources of effort complementarity in teams. First, workers' efforts in stores with two or more workers per shift can be complementary if workers specialize across product categories or tasks within the shift (such as beer vs. soft drink, or warehouse management vs. shelf replenishment). For example, being a specialist in a category means that shared spaces/processes (e.g., warehouse order, product ordering) would be done better and generate positive spillover to the other category; likewise, improving tidiness inside the store warehouse via task specialization substantially facilitates shelve replenishment. Second, sequential shifts generate complementarities: the tasks of one shift are easier to execute if the previous shift executed them well. For example, tidiness in the warehouse increases the return to effort in the current shift as well as in the next one; or, handing over the store with partially empty shelves damages the productivity of the next shift.

One of the key performance indicators of replenishment is store productivity as measured in liters per worker. Workers and supervisors may affect productivity through two non-exclusive methods: i) for a given level of sales, higher effort provision results in staffing fewer workers; and/or ii) sales may be increased by increasing product availability on store shelves. Research indicates that both paths are

[4] Highly imperfect subrogates exist. For approximately half of the stores, some manufacturers collect information (e.g., shelf service level, prices, compliance with trade agreements, share of shelves) through specialized market research firms that randomly visit stores once per week. There is consensus in the industry that although it is valuable, this tool exhibits measurement errors.

important (Corsten and Gruen, 2003; Ton and Raman, 2010). This is consistent with our conversations with executives at the firms we analyzed: they track productivity at the store level (e.g., liters per worker) in order to adjust staff allocations every so often, and they also track “shelf service level,” which is equal to “1 – stock outs,” in order to see whether worker effort is translating into available products on shelfs. In section 4.2.1, we explore the extent to which these two mechanisms explain our results.

2.2 Different strategies in the supermarket channel up to the year 2011

We study two beverage firms that have a combined 90% market share in the Chilean beverage market.[5] These firms are fierce competitors and natural performance benchmarks for each other on many dimensions. In terms of product categories sold, Firm 1 has a diversified portfolio of related products (% of volume in parentheses): soft drinks, including carbonated, juices and water (56%); beer (36%); spirits (4%); and wine (4%). Firm 2 has a portfolio that is focused on soft drinks, with the following breakdown: carbonated soft drinks (77%), water (12%), juice (10%), and other (1%). In terms of relative volume sold in supermarkets, these two companies are of equivalent size: in 2011, Firm 1 has a mean monthly sales volume of approximately 55 thousand liters per store (standard deviation of 49 thousand liters), whereas Firm 2 has a mean monthly sales volume of approximately 48 thousand liters per store (standard deviation of 47 thousand liters).

In Table 2, we provide a detailed comparison of the different contract choices made by the firms as of 2011. The contractual differences can be traced to different organizational strategies for the supermarket segment. Firm 1 was the historical leader in beer production and commercialization in Chile. During the 1990s and 2000s, this firm grew primarily through acquisitions in the related product categories of soft drinks, spirits and wine. Since the early- to mid-2000s, Firm 1’s corporate strategy has been to centralize every function except for production and marketing, which are decentralized into product divisions (or strategic business units, SBUs). In sales, a centralized salesforce sells the full product portfolio; however,

[5] The firms’ identities remain anonymous for confidentiality reasons.

given the importance of supermarkets, servicing them remained fully delegated to each product division in 2011. In addition to replenishment (surveyed above), servicing supermarkets also entails maintaining commercial relationships with different chains and stores, requiring the role of key account managers (KAMs).[6]

By delegating supermarket replenishment to its divisions, Firm 1 rendered pay-for-performance and strong monitoring unviable. First, coordination difficulties among the product divisions led to a situation in which there was no aggregate target for the total liter volume in each store, thereby limiting the use of goal-based pay-for-performance which is frequently used in replenishment and other sales-related roles. Thus, workers of Firm 1 had fixed wages in 2011. Second, monitoring was impaired by several of the features of a decentralized structure: i) store supervisors of different business units did not use standardized operational procedures; ii) divisions typically required store supervisors to have dual roles as KAMs, thus limiting supervision time; iii) due to coordination difficulties, division supervisors overlapped and stores ended up having multiple supervisors, which generated conflicting styles and wider spans of control; and iv) at large stores, workers tended to be (nominally) separated by division (usually across beer and soft drinks), whereas in small stores, individual workers had to be “shared” among supervisors from different divisions, creating further conflicts among supervisors.

In contrast, Firm 2 has been Chile’s historical leader in carbonated soft drinks, achieving growth through diversification into closely related beverage categories such as juices, iced tea, bottled water, and energy drinks. This diversification has been executed organically, leading to a functional corporate structure (including replenishment) with no independent product divisions. This enabled strong formal incentives to be put in place, namely pay-for-performance as well as strong monitoring. Store-specific volume goals are feasible, thereby allowing Firm 2 to pay for performance. Firm 2 pays its workers

[6] Commercial relationships entail annual negotiations and monthly enforcement of trade conditions (e.g., price, shelf space, marketing policies, discount agreements, new product introduction) and monthly coordination of marketing and sales activities (e.g., specific promotions, changes in store equipment, seasonal merchandising, information feedback from supermarkets to producers).

according to a non-linear sigmoidal payout curve that uses a store's volume-to-goal ratio as the input and starts paying from around 70% and is capped around 130%. On average, this goal-based variable pay represents approximately 40% of workers' total compensation. In Figure A1 of the online appendix, we display a representative payout curve used by Firm 2. For large stores, this incentive scheme implies that teams at Firm 2 have a group incentive payment scheme; in small stores, the worker receives an individual pay-for-performance scheme.

In terms of monitoring, Firm 2 is stronger. Firm 1 has 32% fewer supervisors (=15/22-1), and because their store supervisors also perform KAM functions that demand approximately half of their available time (Firm 2, on the other hand, has dedicated KAMs), Firm 1 lags behind Firm 2 in actual supervising time at stores by approximately 65%. Moreover, the span of control differed substantially. Firm 2, being centralized, can optimize the allocation of supervisors to stores, so there is no overlap. These differences create an important difference in the span of control: Firm 1's supervisors manage approximately 50 stores, whereas Firm 2's supervisors manage approximately 20 stores. Further compounding the differences, Firm 2 had standardized control procedures, which Firm 1 lacked. Finally, at both firms, the employment agencies performed monitoring on minimal standards such as punctuality and hygiene. If these minimal conditions were not met, around 5% of base wage would be at risk, per discretion of the employment agency. The denial of this bonus depended primarily on the degree/frequency of monitoring; therefore, for the same level of shirking, workers at Firm 2 were more likely to have this 5% taken away due to greater supervision (cf. Nagin et al., 2002). Thus, monitoring in Firm 2 was both stronger and more impactful.

Even though differences in monitoring are large, pay-for-performance has a stronger impact on compensation; therefore, it should exert a stronger impact on effort decisions. Surpassing the volume goal by 20% generates a variable compensation that is equivalent to 1.2 extra fixed wages; but if volume is 20% short of the goal, the variable compensation is equivalent to 0.12 extra fixed wages (see Figure A1 in the online appendix). In contrast, monitoring can probabilistically generate a reduction of 0.05 fixed wages, an effect that is one degree of magnitude smaller than that generated by pay-for-performance.

Table 2. Difference in contracts between Firm 1 and Firm 2 in the year 2011.

Contract choice dimension		Firm 1	Firm 2
Compensation of agents (pay-for-performance)	Average monthly wage of workers	$580 USD, fixed across workers	$640 USD, with high dispersion across workers
	Pay-for-performance?	No, 100% fixed wages	Yes, performance pay related to store volume accounts for 40% of total wage on average. Steep payout curve between 70% and 130% of store volume goal.
Monitoring of agents	Number of supervisors	15 supervisors	22 supervisors
	Location of supervisors' work	50% office work (KAM role + paperwork) 50% monitoring at stores	10% office work (paperwork only) 90% monitoring at stores
	Supervisors' span of control	~ 50 supermarkets	~ 20 supermarkets
	Standardized procedures of control	No	Yes
	Penalty for not meeting basic requirements? (Service provided by the employment agencies)	Yes, a small portion (approximately 5% of the base wage) could be negated if workers do not perform minimum effort (attendance, hygiene, basic etiquette, uniform, etc.).	Yes, same as Firm 1. However, given the higher frequency of visits to stores by supervisors, losing this portion of base wage is more likely if minimal effort is not exerted.

2.3 Changes after 2011

After 2011, Firm 1 changed the way it organized its supermarket operations, including their contracting practices with their replenishment workers. The firm centralized its supermarket operations: replenishment, instead of being delegated to the product divisions, became a part of the centralized sales unit that Firm 1 already had in place for other sales channels. Contractually, they mimicked the system of Firm 2 almost identically (see Table 2). This meant establishing a physical and managerial infrastructure to manage the new centralized unit, as well as the migration of workforce and contracts from the divisions to the new unit.

This change was made because the firm realized at the end of 2011 that they had a competitive disadvantage in their supermarket operations stemming from their contracting practices (two of the authors were involved in this benchmarking/consulting effort). Also, the centralized sales unit had matured/consolidated enough from their creation in the mid-2000s so that it was well positioned to expand and absorb the supermarket operations.

The timing of the change was as follows. In 2012 the change was designed; 2013 was the first year of piloting where key details were ironed out (e.g., coordination with SBUs, selection of KAM, defining the parameters of the pay-for-performance system, "administrative migration" of workers from SBUs to centralized sales unit). In 2014, the new organization and associated contracts were operating as designed.

Firm 2's supermarket operations did not experience any changes in their organizational or contracting practices. The only important change was a relocation of its headquarters, which affected replenishment via the distance to stores which supervisors must travel, because they begin their workdays in team meetings at the headquarters. We address this change in the empirics. (The change in Firm 1 did not produce a change in distances because supervisors' new and old office spaces remained within the same location.)

3 Formal model and predictions

In this section, we introduce a simple and general model that incorporates the key characteristics and dynamics at play in our setting. We consider a set of employees $i \in \{1,2,\dots N\}$ (agents). When $i = 1$, there is only individual production; when $i > 1$, there is team production. We assume the principal (manager) and the agents are risk-neutral. We assume that an agent's contribution to the collective output, denoted y_i, is a function of the employees' efforts e_i, a stochastic noise term $\varepsilon_i \in N(0, \sigma^2)$ and, drawing on Itoh (1991), the interdependence b between the agents' efforts (which captures how much one agent's contribution influences the other agent's contribution). For simplicity's sake, b affects all agents' tasks the same way and the problem is defined when $b < 1$. The agents' effort decisions impact each agent's performance as follows:

$$y_i = e_i + \varepsilon_i + b \sum_{j \neq i} (e_j + \varepsilon_j)$$

As discussed above, this interdependence may arise across shifts (e.g., warehouse tidiness in the morning shift helps the afternoon shift) but also within a shift (e.g., worker specialization across categories). If $b > 0$ we have the case of effort complementarity. Total performance at the store level is:

$$y_s = y_1 + y_2 + \cdots + y_N$$

Although contributions y_i are prohibitively costly to measure by the principal, it is easy to observe total output at the store level y_s; therefore, it can only place formal pay-for-performance incentives based on y_s. This incentive takes the form of a rate k_s per unit of physical output at the store level (this can be standardized based on an output goal, as in our setting, without changing our results). In a team production setting, each agent receives $\frac{k_s}{N}$.

To further mirror our empirical setting, we assume that the principal can incentivize the agents through a base wage $w^m(e_i)$ which is increasing in effort (see the last row of Table 2). The strength of the formal incentives is captured by the superscript m: if m = "strong" then pay-for-performance ($k_s > 0$) and strong monitoring are present; if m = "weak" then there is no pay-for-performance ($k_s = 0$) and monitoring is weak. We further assume that the strength of monitoring affects how effort translates into the base wage; in particular, we assume $\frac{\partial w^{strong}(e_i)}{\partial e_i} > \frac{\partial w^{weak}(e_i)}{\partial e_i} > 0$ to capture the fact that strong monitoring can more effectively deter shirking.[7] The function $w^m(e_i)$ does not depend on N because supervisors distribute their monitoring time such that all workers receive similar monitoring attention.

[7] This assumption for $w^m(e_i)$ could be obtained from the prevalent practice for modeling monitoring (e.g., Dickinson and Villeval, 2008). Usually, the wage depends on the likelihood of "being caught not exerting minimum effort," as per $w^m(e_i) = W - (W - w) \times \emptyset(e_i, m)$ where W is the baseline wage; w is the penalty wage that the worker receives if they are caught (in the case of our setting, $w = 0.95W$); $\emptyset(e_i, m)$ is the likelihood of being caught, which displays the following plausible properties: $0 \leq$

Employees incur a quadratic cost $c(e_i) = \frac{e_i^2}{2}$ when they exert effort (as it is common in the literature, e.g., Siemsen et. al, 2007).

We add social incentives to the model, which, following Ashraf and Bandiera (2018), corresponds to "any factor that a) affects the marginal benefit or marginal cost of effort and b) stems from interactions with others" (Ashraf and Bandiera, 2018). We model this as a general function $S^m(e_i, e_{-i}, N)$ that includes the employee's own effort e_i, the other employees' effort e_{-i}, and the number of employees N. This function increases worker utility and is such that $\frac{\partial S^m(e_i,e_{-i})}{\partial e_i} > 0$ and $S^m(e_i, e_{-i}, N = 1) = 0$; this means that effort of worker i increases its marginal benefit/utility via $S^m(.)$ and that social incentives occur only in teams.[8] Although we are agnostic about the form that $S^m(.)$ can take, in section 4.2.4 we propose a function that can make sense of the results we will uncover (see Ashraf and Bandiera, 2018, for a fine summary of the generic forms $S^m(.)$ takes). We allow the strength of the social incentives to depend on the strength of formal incentives m, which can be "strong" or "weak" such that $\frac{\partial S^{weak}(e_i,e_{-i})}{\partial e_i}$ may be larger or smaller than $\frac{\partial S^{strong}(e_i,e_{-i})}{\partial e_i}$. Although the literature tends to show that social incentives are more effective under weak incentives (Bowles and Polania-Reyes, 2012; Ashraf and Bandiera, 2018; Herbs and Mas, 2015), our model does not preclude the opposite.

$\emptyset(e_i, m) \leq 1$; $\frac{\partial \emptyset(e_i,m)}{\partial e_i} < 0$; $\emptyset(e_i, m = weak) < \emptyset(e_i, m = strong)$; and $\frac{\partial \emptyset(e_i,m=strong)}{\partial e_i} < \frac{\partial \emptyset(e_i,m=weak)}{\partial e_i} < 0$. This formulation yields $\frac{\partial w^{strong}(e_i)}{\partial e_i} > \frac{\partial w^{weak}(e_i)}{\partial e_i} > 0$.

[8] These two assumptions are consistent with our setting. However, in general, social incentives may occur in individual production (i.e., $S^m(e_i, e_{-i}, N = 1) \neq 0$) and $\frac{\partial S^m(e_i,e_{-i})}{\partial e_i}$ could be negative.

Regarding the assumption and assuming $S^m(e_i, e_{-i}, N = 1) = 0$, there are three types of social incentives that may occur in individual production: relational contracting (e.g., Gibbons, 2005), social comparisons (e.g., Obloj and Zenger, 2017) and organizational identity (Akerlof and Kranton, 2005). In our setting we discard these because: i) worker rotation is high, reducing the importance of repeated interactions and thus, relational contracting; ii) individual workers are assigned to specific stores, which reduces observability of others' income/effort and the importance of social comparison (furthermore, goal-based pay-for-performance reduces potential input-based unfairness due to allocation to "better" or "worse" stores); and iii) in general, workers from employment/staffing agencies (e.g., Adecco, Randstad) are transactional and display low identification towards the agency or the agency's clients, reducing the importance of organizational identity (but not group/team identity, which may still occur in large stores).

Regarding $\frac{\partial S^m(e_i,e_{-i})}{\partial e_i}$, this derivative could be negative when relative incentives are used and when social comparisons are present (Ashraf and Bandiera, 2018). Both cases require that outcome be separable among team members, which is not feasible in our setting.

The worker's utility and the firm's profit at the store level are, respectively:

$$U_i = \frac{k_s}{N} y_s + w^m(e_i) - c(e_i) + S^m(e_i, e_{-i}, N)$$

$$\Pi = (1 - k_s\,) y_s - \sum_i w^m(e_i)$$

In a team of symmetric agents, at equilibrium, each agent provides an optimal productivity effort equal to:

$$e_i{}^* = \frac{1 + b(N-1)}{N} k_s + \frac{\partial w^m(e_i)}{\partial e_i} + \frac{\partial S^m(e_i, e_{-i}, N)}{\partial e_i}$$

This expression shows that although social incentives always increase effort in teams, effort complementarity ($\frac{b(N-1)}{N}$) and free-riding ($1/N$) matter only under strong incentives, because weak incentives (i.e., $k_s = 0$) do not make these two elements a part of workers' effort decisions. Also, the expression makes clear that free-riding in teams is counteracted by effort complementarity, social incentives and monitoring: the first term of e_i^* is increasing in b, the second and third terms increase effort, and the magnitude of this increase depends on the strength of incentives m.

Using this expression, we can easily find and compare the four different effort levels A, B, C and D depicted in Table 1 by modifying the production technology, namely team ($N > 1$) or individual ($N = 1$), as well as the strength of incentives, namely weak incentives (s = "weak"; $k = 0$) or strong incentives (s = "strong"; $k > 0$). We introduce a series of predictions around these four effort levels that we will empirically test in the next section. In the online appendix we provide the proofs of these predictions.

Prediction 1 (A vs. B): *In individual production, weak incentives generate lower effort in comparison to strong incentives. Thus, A < B.*

Prediction 2 (C vs. D): *In team production, the difference between effort levels C and D can be small, and thus the advantage of strong incentives can be less powerful for teams than for individuals (i.e., B – A > D – C), under two conditions:*

1) If social incentives are symmetric (i.e., the same between strong and weak incentives), and b *does not converge to 1, then free-riding reduces D as N increases and D converges to C. However, a gap will remain due to monitoring being stronger under strong incentives (i.e., C < D).*

2) If social incentives are asymmetric and sufficiently greater under weak incentives than under strong incentives, C can be equal to, or even larger than D. The higher the value of b*, the higher the asymmetry in social incentives must be.*

Prediction 3 (A vs. C): *Under weak incentives, teams exert more effort than individuals due to social incentives. Thus, A < C.*

Prediction 4 (B vs. D): *Under strong incentives, if the strength of social incentives is large enough, teams will exert more effort than individuals. Thus, D > B.*

Prediction 1 is straightforward. Regarding Prediction 2, imagine that one obtains C < D and that the difference is small (e.g., Friebel et al., 2017). Part 1 of this prediction indicates that this difference could be generated by the combination of symmetric social incentives, low effort complementarity and free-riding; in this case, *the small difference between C and D derives from strongly incentivized teams reducing their effort; that is, "D goes down."* Part 2 of the prediction indicates another possibility, namely that effort complementarity is strong, curtailing free-riding, and thus, the result is produced because the effort of *weakly incentivized teams gets strongly lifted due to social incentives; that is, "C goes up."* Knowing whether effort complementarity is strong or weak is crucial for disentangling these two possibilities. Foreshadowing our empirical analysis, the case of $N = 2$ will be crucial in our case: with two workers, social incentives are negligible (due to non-overlapping shifts) but effort complementarities across shifts are present. This will allow us to measure b.

Testing Prediction 3 will also contribute to documenting the presence of social incentives in teams. Given that under weak incentives there is no impact of free-riding and effort complementarity on teams, and that monitoring does not depend on N, C – A provides an estimate of the size of social incentives for weak incentives (i.e., $\frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}$).

Finally, note that by analyzing (B – A) – (D – C) (or its equivalent, (C – A) – (D – B)), one eliminates the influence of monitoring in the analysis: it is easy to show that (B –A) – (D – C) = $\left(1-\frac{1+b(N-1)}{N}\right)k_s + \left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}\right)$, and is thus independent of $w^m(e_i)$.

4 Empirical analysis

4.1 Data

Our dataset covers 175 supermarket stores located in Santiago de Chile. For each firm, for every month between January 2011 and August 2011, and between January 2014 and June 2015,[9] we gathered the following data in each store: sales volume (in liters) separated by product category, number of workers, GPS coordinates, square meters, county where the store is located (43 in total), the supermarket chain (six), and the human resource agencies that provide services to each firm for each store (two for Firm 1 and three for Firm 2). In Figure 1, we display the productivity distribution at the store level, measured as liters per worker, for both firms in two periods of time: 2011 and 2014/2015. The distribution of Firm 1 moves to the right in 2014/2015, consistent with the expectation of strong incentives increasing productivity. In Figure A2 of the online appendix, we display the average of productivity over time for both firms; whereas Firm 2's productivity is quite stable over time, Firm 1 lags behind Firm 2 in 2011 but catches up in 2014 and 2015.

[9] There is no data in 2012 and 2013 for two reasons. First, data for Firm 2 was collected manually in 2011 as part of a consulting project that involved two of the authors. Firm 2 started systematically collecting data on replenishment workforce allocation to stores at the end of 2013. Second, Firm 1 designed the organizational change in 2012 and the first year of implementation was 2013, during which many adjustments were made. From 2014 onwards, the change was operating as expected.

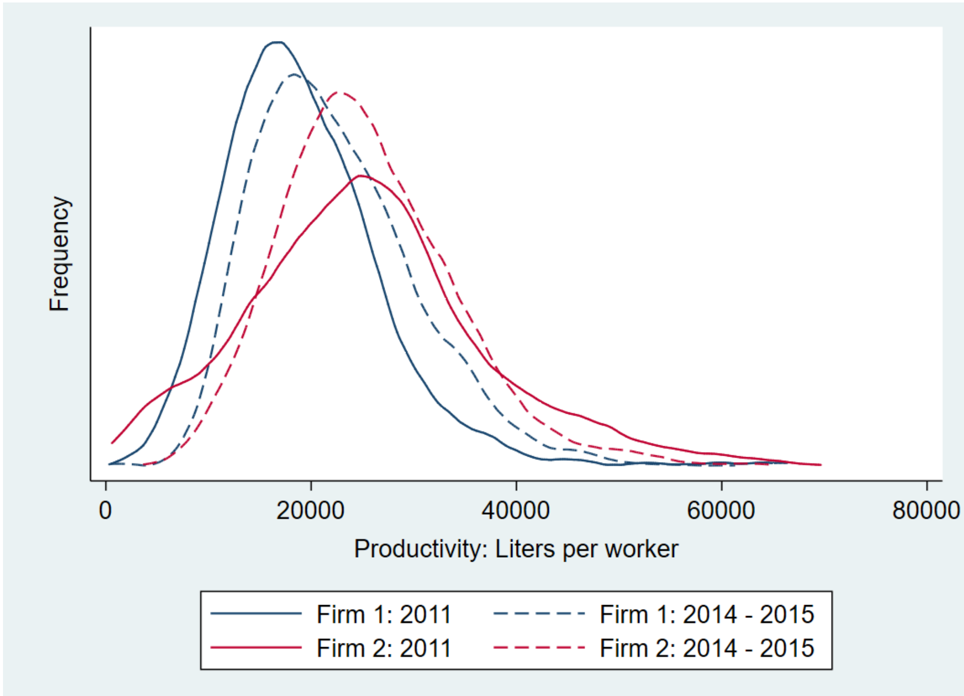

Figure 1. Kernel density of store-level productivity

4.2 Econometric analysis

To analyze the impact of contracting in the productivity of firm j in the store i in the period t, we use the following model:

$$\text{Ln(Productivity)}_{ijt} = b_1 + b_2 \text{ x Weak\_incentives}_{ijt} + u_{ij} + v_t + e_{ijt} \quad (1)$$

"Weak incentives" is a dummy variable that takes the value of 1 for Firm 1 in the year 2011, and 0 otherwise. We add store-firm fixed effects u_{ij} and period-fixed effects v_t (26 periods in total). The coefficient b_2 captures the average impact of having weak incentives (i.e., fixed wage and low monitoring) across individual and team production. For example, if $b_2 = -0.1$, it means that weak incentives have a productivity disadvantage of 10%. To analyze Predictions 1 and 2, we use the following two models:

$$\text{Ln(Productivity)}_{ijt} = \beta_1 + \beta_2 \text{ x Weak\_incentives}_{ijt} + \beta_3 \text{ x Weak\_incentives}_{ijt} \text{ x Store\_size}_i + u_{ij} + v_t + e_{ijt} \quad (2)$$

$$\text{Ln(Productivity)}_{ijt} = \alpha_1 + \sum_q (\alpha_q \text{ x Weak\_incentives}_{ijt} \text{ x q\_workers}_{ijt}) + u_{ij} + v_t + e_{ijt} \quad (3)$$

In model (2), the variable “Store size” is measured as the natural logarithm of the square meters of the store (results do not change if logarithm is not used). At the 10th percentile of this variable, where ln(square meters) ~ 6.5, there is 1 worker per store; at the 90th percentile, where ln(square meters) ~ 9, there are 4 to 5 workers. In Figure 2, we display a scatterplot of store size and number of workers.

Model (3) is used to exploit the discrete changes in the number of workers in the site. In this model, we use the average number of firm–store workers in 2011 and create the following set of dummy variables: “1_worker” takes the value of 1 if the average number of workers in 2011 is equal to or lower than 1.5; “2_workers” takes the value of 1 if the average is higher than 1.5 and equal to or lower than 2.5; “3_workers” takes the value of 1 if the average is higher than 2.5 and equal to or lower than 3.5; “4_workers” takes the value of 1 if the average is higher than 3.5 and equal to or lower than 4.5; and “5_workers” takes the value of 1 if the average is higher than 4.5. We split the impact of “Weak incentives” across these five dummies.

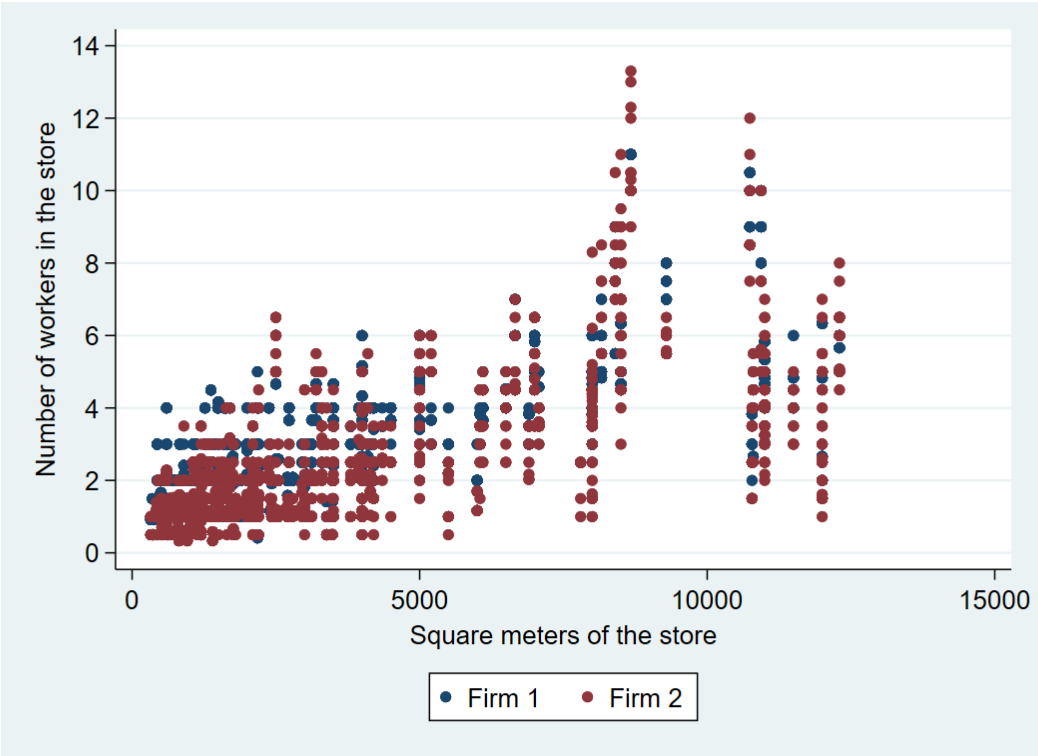

Figure 2. Scatterplot of store size and number of workers

In Table 3, we display the results. Column (1) shows that under weak incentives, workers display a lower productivity of 12.4%, on average, as compared to the productivity under strong incentives (p-value < 0.05). Column (2) shows that this productivity disadvantage is heterogeneous across store size (p-values for the dummy "Weak incentives" and its interaction with size are both low—below 0.01). In a store at the 10^{th} percentile for size, where ln(square meters) is roughly 6.5, the productivity disadvantage is equal to 22% (= - 0.753 + 0.082 x 6.5). In column (3) we find the same result: when there is one worker in the store, the productivity disadvantage is 18.6% (p-value < 0.05). These results mean that under individual production, weak incentives lead to a large penalty in productivity; that is, B > A, as highlighted in Prediction 1. This is aligned with the literature: "most of the field and lab experiments that identify the causal impact of individual bonuses on productivity find a 20–25% increase" (Ashraf and Bandiera, 2018; p. 444).

In contrast, column (2) shows that, in a store at the 90^{th} percentile for size, where ln(square meters) is roughly 9 and there are teams of four or more workers, the productivity disadvantage is 2% (= - 0.753 + 0.082 x 9); this is not statistically different from zero. Column (3) shows that when there are four workers, the disadvantage is 4% (not statistically different from zero), and when there are five or more, it is 0%. These results mean that under team production, weak incentives and strong incentives do not display a difference in productivity; that is, C = D. This result is consistent with the small effect of 3% found in Friebel et al. (2017) and the null effect reported by Delfgaauw et al. (2020 and 2022).

As discussed in Prediction 2, a very low difference (or equality) between C and D may be achieved in two ways. On the one hand, it could be that social incentives are symmetric and effort complementarity is very low, and thus, that free-riding reduces D, which converges to C. On the other hand, if effort complementarity is high, it will strongly counteract free-riding, in which case the asymmetry in social incentives in favor of weak incentives must be large in order to generate D = C. To determine which of the two mechanisms drives the equality of the productivities D and C, we executed three analyses: i) exploit the discrete changes in workers to obtain an estimate of the size of the effort complementarities (parameter b of the model), ii) exploit the discrete changes in workers to evaluate the non-linear impact of free-riding, and iii) estimate Predictions 3 and 4 in order to further assess the presence of asymmetric social incentives in our setting.

First, let us analyze the discrete changes in workers in column (3). The results show that teams of two workers display a 16.7% productivity disadvantage under weak incentives (in comparison to strong incentives) (p-value < 0.01)—a mere 1.9 percentage-point difference from the productivity disadvantage of individuals, which is measured at 18.6%. Using the formulas from the appendix, it is easy to show that this difference of 1.9 percentage points is equal to (B – A) – (D – C) = $\left(1-\frac{1+b(N-1)}{N}\right)k_s + \left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}-\frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}\right)$. Given that a team of two workers has little to no social interactions because shifts are non-overlapping—and thus no access to social incentives—one may assume that social incentives are negligible and obtain $\left(1-\frac{1+b}{2}\right)k_s = 1.9\%$. From column (3) of Table 3, we also know that B – A is equal to 18.6% = $k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i}-\frac{\partial w^{weak}(e_i)}{\partial e_i}\right)$. Assuming that the last term of monitoring asymmetry is not the main driver of strong incentives (see section 2.2 above), these equations imply a large k_s and, as a result, a large b. For example, if k_s equals 10% (and thus, monitoring asymmetry equals 8.6%), 12% (and thus, monitoring asymmetry equals 6.6%), or 14% (and thus, monitoring asymmetry equals 4.6%), then b equals 0.62, 0.68 and 0.73, respectively. Large values of b are expected when two workers work in sequential shifts. A value of b = 0.7 means that if workers of the preceding shift were absent, the output of the worker in the next shift is reduced by 41%, according to our model.[10] This is intuitive in our setting: output in the afternoon shift will suffer if the shift starts with nearly empty shelves, because it takes time to restock the shelves to an adequate level. Note that this is a minimum estimation of b, because it considers only the complementarity that occurs between shifts, and not that

[10] Output in 2nd shift is equal to $y_2 = e_1 + 0.7\, e_2$. Given that agents are symmetrical, $e_1 = e_2 = e$; thus, $y_2 = 1.7e$ if the morning shift is present and $y_2 = 1e$ if the morning shift is not. The percentage difference is 1-1/1.7 = 41%.

which occurs within a shift. For example, due to task specialization[11] (if we could add this second type, b would be even higher and our results in terms of asymmetric social incentives would be stronger). This analysis shows that effort complementarity is large, and will thus effectively counteract free-riding. This suggests that instead of D going down, the social incentive mechanism is the most likely driver in generating C = D by "lifting" C.

The second analysis we perform is to explore how changing the number of workers impacts productivity. If free-riding were driving D = C, and given the non-linear impact of N on optimal effort, we should observe that the decrease in D – C accompanying an increase from two to three workers should be twice as large as the decrease in D – C when going from three to four workers. However, we find that the decrease is (quasi) linear. With three workers, D – C corresponds to 11.6% of productivity (a change of 5.1 percentage points from two workers), and with four workers the gap is 4.3% (a change of 7.3 percentage points from three workers). This (quasi) linear change is inconsistent with a free-riding-only explanation, but it is consistent with the notion of social incentives becoming relevant in a step-by-step fashion. This gradual addition of social incentives is plausible in our setting. When increasing from two to three workers, social incentives can be present in only one of the two shifts; therefore, social influence is exerting only half of its potential in such a case. However, when increasing from three to four workers—meaning that both shifts have two workers—then we find that the impact reduction in the D – C is approximately doubled, which is consistent with the influence of social incentives during both shifts. Again, this argues in favor of social incentives lifting C rather than free-riding alone reducing D.

The third analysis is to leverage Predictions 3 and 4 to obtain a measure of social incentives under weak and strong incentives, respectively. As shown in the appendix, Predictions 3 and 4 show that C > A and D > B are generated exclusively by social incentives. In order to estimate both predictions, we use the following model:

[11] Of the two types of task specialization—by physical area (e.g., warehouse vs. aisles/shelves) or by product category (beer vs. soft drink)—the latter type was already incorporated by Firm 1 under weak incentives, and thus, the impact of this type of complementarity under the new contract is not relevant, thereby reducing the under-measurement problem of b.

$$\text{Ln(Productivity)}_{ijt} = \sigma_1 + \sigma_2 \times \text{Weak\_incentives}_{ijt} + \sigma_3 \times \text{Weak\_incentives}_{ijt} \times \text{Store\_size}_i + \sigma_4 \times \text{Store\_size}_i + u_j + v_t + e_{ijt} \quad (4)$$

In this model, we control for firm-fixed affects only in order to allow for an estimate of the baseline impact of size. We estimate this model in column 5 of Table 3 and find that σ_2, σ_3 and σ_4 all display low p-values (< 0.01). The results show that, under weak incentives, the elasticity of size to productivity is equal to 0.161 (= 0.068 + 0.092), which means that going from the 10th to the 90th percentile in store size (i.e., a measure of C – A) translates into an increase in productivity of 40.3% (0.161 x (9 – 6.5)), which is statistically significant. This provides evidence in favor of Prediction 3 and a "direct" measure of social incentives as $C - A = \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}$. Under strong incentives, we find that the elasticity of 0.092 (p-value < 0.01), meaning that going from the 10th to the 90th percentile in store size translates into an increase in productivity of 23% (0.095 x (9 – 6.5)). Given that 23% = $D - B = k_s\left(\frac{1+b(N-1)}{N} - 1\right) + \frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}$, and we know that k_s and b are large, it is simple enough to do the algebra and find that $\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}$ amounts for at least 23% (but closer to 23% as b increases). This provides evidence in favor of Predictions 3 and 4, suggesting that social incentives are present, large and asymmetrical (with a difference of approximately 17% between weak and strong incentives).[12]

Overall, our results suggest that: i) strong incentives improve the productivity of individual workers, and that ii) in teams, effort complementarity is strong, curtailing free-riding, and that social incentives are

[12] Of course, these conclusions regarding C – A and D – B require that there be no other mechanism generating higher productivity with size than those mechanisms included in our model. Our conversations with firm executives suggested one such potential mechanism, namely economies of scale in avoiding the need for workers to "travel," or commute between stores. In small stores, where fewer than two full-time workers are needed (i.e., one per shift), workers are assigned to two or three other nearby stores, obligating them to "rotate" between stores (this generates the decimal numbers in Figure 2). These traveling times are a "deadweight" loss in terms of productivity, which is avoided by larger stores where workers are fully assigned to one store (in these stores, one half or one quarter of store workers are added as extra capacity for peaks—for example, on weekends—and they do not move between stores). We estimated that, at most, these travels require 1 hour of their time in a 9-hour workday, leading to an 11% (= 1/9) reduction in productivity. This means that the estimation of social incentives from B – A and D – B—that is, 40% under weak incentives and "at least 23%" under strong incentives—must be adjusted to roughly 29% and "at least 12%," respectively.

higher under weak incentives; as a result, weakly incentivized teams are just as productive as strongly incentivized teams.

Table 3. Impact of weak incentives on productivity across individual and team production

	Dependent variable:							
	Ln (Productivity)						Ln (Volume)	Ln (Workers)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weak incentives	-0.124 (0.049)	-0.753 (0.212)			-0.637 (0.203)		-0.354 (0.156)	0.399 (0.190)
Weak incentives x Store size		0.082 (0.027)			0.068 (0.026)		0.047 (0.019)	-0.035 (0.023)
Weak incentives x 1_worker			-0.186 (0.079)					
Weak incentives x 2_workers			-0.167 (0.063)					
Weak incentives x 3_workers			-0.116 (0.115)					
Weak incentives x 4_workers			-0.043 (0.077)					
Weak incentives x 5+_workers			0.009 (0.059)					
Store size				0.102 (0.013)	0.092 (0.013)			
Shelf service level						0.313 (0.095)		
Period FE?	YES	YES	YES	YES	YES	YES	YES	YES
Firm x Store FE?	YES	YES	YES	NO	NO	YES	YES	YES
Firm FE?	NO	NO	NO	YES	YES	NO	NO	NO
Constant?	YES	YES	YES	YES	YES	YES	YES	YES
R-square	54.2%	54.5%	54.5%	19.2%	19.9%	58.4%	92.1%	89.9%
Observations	9,008	9,008	9,008	9,008	9.008	4,550	9,008	9,008
* p-value<0.1; ** p-value<0.05; *** p-value<0.01. Cluster-robust standard errors, clustered at the firm-store level.								

Our results and model allow us to decompose the overall effect of B – C across the different generic mechanisms of free-riding, effort complementarity, and social incentives. Assuming that b = 0.7, N = 5 and that $\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}$ is equal to 2%,[13] the result B – C = 0% is driven by i) a 19 percentage-point (pp) increase in D due to formal incentives (i.e., 2 pp due to monitoring assumption and 17 pp due

[13] In our setting, we do not observe the impact of asymmetry in monitoring. We assume 2% because pay-for-performance generates an impact on compensation that is one order of magnitude larger than monitoring; thus, it should be more important in driving workers' effort choices and productivity (see the discussion in section 2.2 of the paper above).

to pay-for-performance (B – A = 19 pp = $k_s + \frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i} \Rightarrow k_s$ = 17 pp); ii) a 13.6 pp reduction in D due to free-riding (= 17 x (1 - 1 / 5)); iii) a 12.9 pp increase in D due to effort complementarity (= 17 x {1 + 0.7 x (5 – 1)}/5); and iv) a residual effect of "asymmetric social incentives" that increases C (relative to D) by 18.3 pp (= {2 + 17 – 13.6 + 12.9} – 0). Note that changing the assumption $\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}$ would change the size of free-riding and effort complementarity, but not that of asymmetry in social incentives (this is made clear by examining the expression for (B – A) – (D – C)). In Figure 3, we present this calculation schematically.

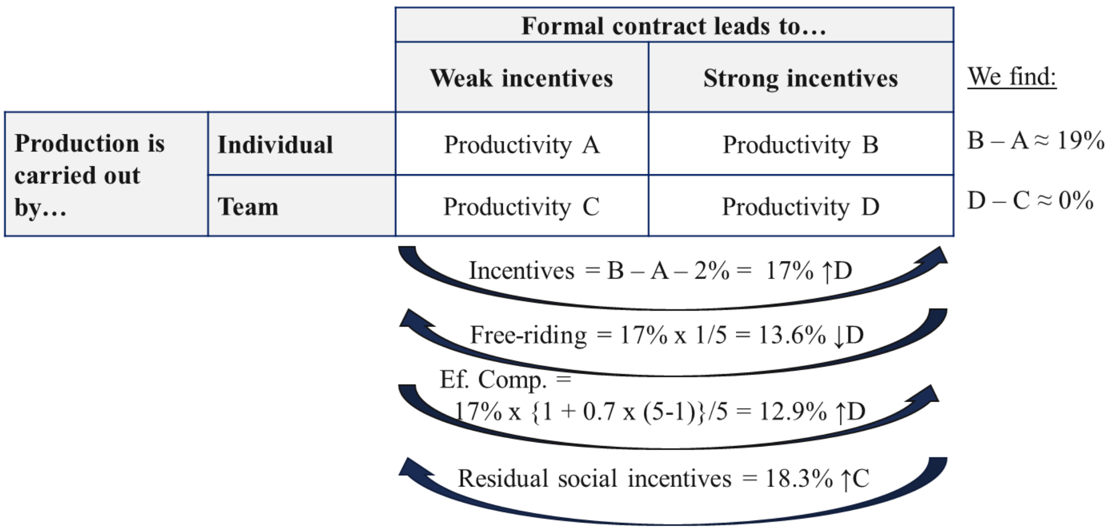

Figure 3. Decomposition of the mechanisms underlying D – C = 0.
(Example using N = 5, and assuming $\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}$ = 2%)

4.2.1 Channels: increase in sales vs. reduction in number of workers

The productivity changes we documented can be generated via two generic channels: 1) an increase due to the availability of products on the shelves, which drives an increase in sales volume—the "sales channel"—and 2) an increase in worker productivity that is detected by the firm, which then reduces the number of workers in the store (e.g., eliminating part-time support workers on weekends)—the "allocation channel." Both channels involve workers' response to incentives. However, whereas the sales channel is a direct channel that provides a clear indication of workers' behavioral response to incentives (as required by our model), in the allocation channel it is not possible to rule out the possibility that the firm "imposes" a productivity level by fiat, which would not necessarily elicit a reasoned behavioral response by workers to the contract change (instead, they comply with commands). In columns (7) and (8) of Table 3, we explore the relative weight of these channels. In column (7), the coefficients for the dummy "Weak incentives" and its interaction with "Store size" display low p-values and show that, relative to strong incentives, weak incentives are associated with a 5% lower volume (= -0.354 + 0.047 x 6.5) under individual production (i.e., 10th percentile of store size), compared to a 7% (= -0.354 + 0.047 x 9) higher volume under team production (i.e., 90th percentile of store size). This means that the "sales channel" accounts for 12% (= 7% - (-5%)) of the 20% productivity differential between individual and team production that we uncovered in column (2). In contrast, in column (8), where we study Ln(workers) as the dependent variable, the interaction term is less precisely estimated. A joint *t*-test on ∂Ln(workers)/∂weak_incentives displays a low p-value (<0.01). Under weak incentives, a small store size is associated with a 17% higher number of workers (= 0.399 – 0.035 x 6.5), and a large store size is associated with an 8% higher number of workers (= 0.399 – 0.035 x 9). This means that the "allocation channel" accounts for 9% of the 20% differential obtained in column (2). These results suggest that both channels are playing a role, but the evidence is slightly stronger in favor of the "sales channel."

Based on these results, we ask whether the 12% differential of the "sales channel" is due to a reduction in stock-outs. In order to find an answer, we obtained information on "shelf service level," measured as "1 – percentage of SKUs with stock-out." Firm 2 measures stock-outs for a subset of all its SKUs in ~95 stores across all years of the sample; Firm 1 does the same in ~120 stores, but only in the years 2014 and 2015 (in 2011, Firm 1 did not measure stock-outs). The variable "shelf service level," when pooled across the two firms, has a mean of 80%, a standard deviation of 13%, and its 10th and 90th percentiles are equal to 63% and 95%, respectively. Although the lack of 2011 data for Firm 1 hinders an evaluation of the impact of formal incentives on "shelf service level," in column (6) of Table 3 we estimate the association of this variable with productivity. We find that a 10% increase in "shelf service level” increases productivity by 3.1% (p-value < 0.01); supplementary analysis shows that this is entirely driven by sales, not number of workers. If “shelf service level” rises from the 10th to the 90th percentile, sales and productivity increase by 10 percentage points (= 3.1% x (95% - 63%)). Thus, if higher worker effort in non-incentivized teams translates into such a change in “shelf service level,” that would account for a majority of the 12% effect generated via the “sales channel.”[14] Therefore, we can conclude that improved “shelf service level” can explain the increase in sales and productivity we observed.

4.2.2 Identification

We address several concerns regarding the endogeneity of our analysis, namely: time-variant unobservables; matching between contract and production technology (team vs. individuals); exogeneity of store size; and worker sorting into production technology (team v/s individuals), firm, and/or incentive scheme. Although these analyses cannot eliminate all concerns, they increase the confidence in a causal interpretation of our results.

Time-variant unobservables. There may have been unobserved changes between 2011 and 2014-2015 that drive the results. Our data allows us to address three time-variant variables at the firm-store level: i) Firm 2 changed its headquarters in the interim between 2011 and 2014, which increased the average distance supervisors had to travel to reach Firm 2 stores; ii) percentage of product categories (i.e., for Firm 1, percentage of volume sold in beer, soft drink, spirits, and wine; for Firm 2, percentage of volume sold in soft drink, water, juices, and other); iii) percentage of volume in “returnable” sales (i.e., both firms have a policy allowing customers to return empty bottles in exchange for a discount on subsequent purchases). These three variables relate to monitoring intensity, product strategy, and reverse logistics, respectively. In Table A1, we show that our results do not change when controlling for these time-variant unobservables.

[14] The remaining 2% can be explained by reduced stock-outs in the SKUs that are not tracked by the firms—or, as we discovered in our field interviews with workers, by creative approaches to utilizing extra space in stores (e.g., by placing soft drinks close to tills) that are not tracked by the firms in the shelf-service measures.

An additional way to tackle this concern is to estimate the bounds on omitted variable bias using the test proposed by Oster (2019). Using column (7) of Table 3 and assuming a maximum R^2 = 0.95 (i.e., we allow productivity to have a small random element), we find that the delta is 1.38 for the interaction between "weak incentives" and "store size." This means that selection on unobservables needs to be 1.38 times the selection on observables in order to overthrow our results. This is beyond the suggested threshold of 1 by Oster (2019), and indicates that time-variant unobservables may not be a significant concern.

Match between production technology and contract choice. A key feature of our empirical analysis is that contract choice and production technology (individual vs. team production) are not endogenously matched. Because Firms 1 and 2 choose one type of contract across stores, variance in store size—and, thus, in production technology—is exogenous to contract choice. This feature in our data is unique in comparison to prior studies, which have instead resorted to complex econometric techniques to clean the estimates of endogenous matching between contract type and production technology (e.g., triple differences analysis by Burgess et al., 2010).

Exogeneity of store size. Even if the "weak incentives" dummy is exogenous, "store size" must be exogenous in order to causally identify the interaction term between them. However, store size is not random and may correlate with other store characteristics that drive the results. We mitigate this concern by controlling for the interaction of "weak incentives" with a large set of store characteristics we collected in 2011 through a survey of Firm 2's supervisors. In Table A3 of the online appendix, we detail these variables and show that our results are robust to their inclusion as a control.

Worker sorting. Our results could be affected if workers systematically sort into firm characteristics (including incentive strength) and store size. In particular, the results could be driven by a selection effect rather than a treatment effect.

Regarding *sorting into firm characteristics (including incentive strength),* we first indicate that in our context such sorting is unlikely, and then provide econometric evidence that consistent with sorting. From our interviews with firm and employment agency executives, we learned that these workers' first-order decision is to be employed by an employment agency (e.g., Adecco, Manpower), superseding the decision about the type of activity they perform within the employment agency or which agency client they work for. Once a worker enters an agency, the agency exercises discretion in assigning workers to supermarkets, retailers, logistics, or any other economic sector or task. Of course, if a worker has some experience in supermarket replenishment, the agency will tend to allocate them to replenishment activities (but not necessarily to a particular firm). Overall, this reduces the concern of worker self-selection. To further reduce this concern, we exploit the fact that, although agencies work usually work for only one firm in our setting, there was one exception. In the years 2014 and 2015, Firm 2 used one of the agencies that Firm 1 also used throughout 2011–2015. This overlap of firms within a single agency would facilitate worker sorting into firm characteristics because it is far easier for workers to *move across client firms within the same agency* than to shift agencies altogether in order to shift firms. In Table A3 of the online appendix, however, we show that the coefficients of "weak incentives" and their interactions with "store size" are not affected by the presence of a common agency. This undermines the plausibility of a sorting explanation, which would suggest that the coefficient would be boosted by the common agency and the facilitation of self-selection.

Regarding *sorting into size*, the evidence we gathered argues against it. For Firm 2 workers in 2011, we collected several observables that allow us to analyze whether—for this subsample—there is any difference in the characteristics of workers in large stores vs. small stores. We gained access to five variables that may signal worker "type": age, tenure in replenishment for Firm 2, marital status, percentage of wage attributable to overtime, and percentage of wages requested to be paid in advance (i.e., on the 15th day of the month instead of the 30th). In Table A4 of the online appendix, we demonstrate that for these five characteristics, there is no statistically significant difference across store size.

4.2.3 Tracing social incentives

So far, we have not directly observed social incentives. Instead, we have inferred their existence as a residual that is obtained after accounting for the remaining generic drivers that are predicted to affect worker effort in teams according to the literature (i.e., free-riding, effort complementarity, monitoring). In this section, we supplement this inference by providing more direct evidence for the existence of social incentives.

Researchers in many disciplines have posited and provided substantial evidence in favor of the idea that pro-sociality within a group or team is triggered by competition and threats coming from rival groups and teams (e.g., Bauer et al., 2016; Gelfand et al., 2009; Francois et al., 2018; Bohm and Rockenbach, 2012; Bowles, 2006). To test this idea, we analyzed the reaction of a focal team to an exogenous increase in the productivity of the co-located rival team, and we assessed whether this reaction depended on the strength of incentives. Pro-social effort should be triggered when facing more productive rivals, thus increasing the effort and the productivity of the team. Given our theory and previous findings, this response to rivals is predicted to be stronger in teams with weak incentives, and weaker in those with strong incentives.

We model the productivity of Firm 1 as a function of its rival's productivity:

$$Ln(Productivity)_{i1t} = \delta_1 + \delta_2 \text{ x } Ln(Productivity\_rival)_{i1t} + u_{i1} + v_t + e_{i1t} \quad (4)$$

In this equation, the coefficient δ_2 captures an elasticity: if the worker(s) in Firm 2 increase their productivity by 10%, the worker(s) of the Firm 1 increase their productivity by 10% x δ_2. We label this elasticity as "reaction to competitor" (RtC).

The coefficient δ_2 is subject to endogeneity bias of the same nature as the issues involved in the estimation of peer effects (Manski, 1993). To correct this problem, we use an instrumental variable technique recommended by Soetevent (2006). The instrument we use for the productivity of Firm 2 is the distance from the store to Firm 2's headquarters, measured as the natural logarithm of the geodesic distance (in km). Firm 2 relocated to a new headquarters at the start of 2013, which permits the inclusion of store-fixed effects. This is a desirable instrument because, conditional on store-fixed effects (which capture the time-invariant distance from the store to Firm 1's headquarters), the distance from the store to Firm 2 headquarters should impact the effort/productivity of Firm 2 but not that of Firm 1. In particular, according to prior research on the role of distance in moral hazard settings (Brickley and Dark, 1987; Jensen et al., 2020; Presbitero and Rabellotti, 2014), increased distance should negatively impact productivity. In column (1) of Table 4, we exploit Firm 2's variance in distance to identify the effect on both its own productivity as well as that of Firm 1. As expected, we find that the distance of Firm 2 reduces its own productivity (i.e., the coefficient of Ln(Distance) is negative and has a low p-value) but not that of Firm 1 (i.e., the coefficient of Ln(Distance_rival) is statistically indistinguishable from zero).

To analyze the heterogeneity of RtC, we use the following model:

$$\text{Ln(Productivity)}_{i1t} = \delta_1 + \delta_2 \times \text{Ln(Productivity\_rival)}_{i1t} + \delta_3 \times \text{Ln(Productivity\_rival)}_{i1t} \times \text{Store\_size}_i + \delta_4 \times \text{Weak\_incentives}_{i1t} + \delta_5 \times \text{Weak\_incentives}_{i1t} \times \text{Store\_size}_i + \delta_6 \times \text{Ln(Productivity\_rival)}_{i1t} \times \text{Weak\_incentives}_{i1t} + \delta_7 \times \text{Ln(Productivity\_rival)}_{i1t} \times \text{Weak\_incentives}_{i1t} \times \text{Store\_size}_i + u_{i1} + v_t + e_{i1t} \qquad (5)$$

$$\partial \text{Ln(Productivity)}_{i1t} / \partial \text{Ln(Productivity\_rival)}_{ijt} = \delta_2 + \delta_3 \times \text{Store\_size}_i + \delta_6 \times \text{Weak\_incentives}_{ijt} + \delta_7 \times \text{Weak\_incentives}_{ijt} \times \text{Store\_size}_i \qquad (6)$$

The interactions in equation (5) enable us to study how RtC changes depending on the store size and on the strength of incentives. We follow Wooldridge (2002: 236–237) to correct for endogenous interaction terms in equation (5).[15] The derivative displayed in equation (6) shows that RtC depends on a baseline level (δ_2), store size (δ_3) and weak incentives (δ_6), as well as on the interaction between store size and weak incentives (δ_7). Our theory would suggest that the coefficient δ_7 is larger than 0: if the productivity of the rival increases (in the case of our instrument, caused by a reduction in the distance

[15] This technique is simple. Suppose a model Y = b1 * X1 + b2 * X2 + b3 * X1*X2, where X1 is endogenous and Z is a valid instrument. Wooldridge (2002) shows that the correct IV estimation is as follows: First obtain X1_HAT by predicting X1 from a regression of X1 on Z and X2; and second, compute X1_HAT * X2, which becomes the instrument for X1*X2. We follow the same technique for the triple interaction in equation (5).

between the rival's headquarters and their stores), then teams that are weakly incentivized should display a stronger RtC than teams that are strongly incentivized.

In the Table 4, we show our results. In columns (2) and (3), using OLS and the instrumental variable technique, respectively, we estimate equation (4). We find that there is a positive bias in RtC, and that after correcting for endogeneity, the productivity of Firm 2 does not generate a change in the productivity of Firm 1 (i.e., the coefficient of Ln(Productivity_rival) is positive with a low p-value in column (2) but not statistically different from zero in column (3)). In columns (4), (5) and (6), we explore the heterogeneity of RtC. In all three cases, the derivative $\partial \text{Ln(Productivity)}_{i1t} / \partial \text{Ln(Productivity\_rival)}_{i1t}$ has high levels of statistical significance in a joint *t*-test. In column (5), and maintaining store size at its mean of 7.63, we find that store size does not affect RtC; however, we find that contracting does: under weak contracting, RtC is equal to 17.6% (= -0.249 + 0.382 + 0.006 x 7.63), whereas under strong contracting, RtC is equal to -20.3% (= -0.249 + 0.006 x 7.63). In column (6) we explore the extent to which team or individual production drives the positive result obtained under weak contracting. To interpret this result, we compute RtC for the following four possible scenarios: 1) when store size is low (10^{th} percentile) and incentives are strong, RtC is equal to -0.70 (= -0.880 + 0.027 x 6.5); 2) when store size is low (10^{th} percentile) and incentives are weak, RtC is equal to -0.52 (= -0.880 + 0.027 x 6.5 - 1.957 + 0.329 x 6.5); 3) when store size is high (90^{th} percentile) and incentives are strong, RtC is equal to -0.63 (= -0.880 + 0.027 x 9); and 4) when store size is high (90^{th} percentile) and incentives are weak, RtC is equal to 0.37 (= -0.880 + 0.027 x 9 - 1.957 + 0.329 x 9).

These results lend support to our theory: When a rival team increases its productivity by 10%, a weakly incentivized team increases its own by 3.7%—whereas a strongly incentivized team reduces its productivity by 6.3%. This is also consistent with prior literature positing that pro-social effort is triggered by a strong rival—but only when pro-sociality is not crowded out by strong incentives. It also provides more direct evidence of asymmetric social incentives.

Table 4. Instrumental variable regressions analyzing the reaction to a competitor's productivity increase.

	Ln(Productivity)					
	Two firms	Firm 1				
	(1)	(2)	(3)	(4)	(5)	(6)
Method:	OLS	OLS	Instrumental variables: Ln(Distance_rival) as instrument for ln(Productivity_rival)			
Ln(Distance)	-0.093 (0.052)					
Ln(Distance_rival)	0.015 (0.050)					
Ln(Productivity_rival)		0.190 (0.027)	0.065 (1.098)	0.522 (2.187)	-0.249 (3.388)	-0.880 (3.392)
Ln(Productivity_rival) x Store size				-0.053 (0.137)	0.006 (0.227)	0.027 (0.129)
Ln(Productivity_rival) x Weak incentives					0.382 (0.664)	-1.957 (4.304)
Ln(Productivity_rival) x Weak incentives x Store size						0.329 (0.698)
Weak incentives x Store size						-3.327 (7.188)
Period FE?	YES	YES	YES	YES	YES	YES
Firm x Store FE?	YES	NO	NO	NO	NO	NO
Store FE?	NO	YES	YES	YES	YES	YES
Constant?	YES	YES	YES	YES	YES	YES
R-square	53.45%	59.00%				
Observations	8,812	4,378	4,280	4,280	4,280	4,280

Robust standard errors in parentheses, clustered at the store–firm level. In Model 1, the within-store variance in distance for Firm 1 is used to estimate the coefficient of Ln(Distance) and Ln(Distance_rival). In column 2, Ln(Distance) is dropped because this is store–invariant for Firm 1. In columns 4, 5 and 6, the test of the derivative ∂Ln(Productivity)/∂Ln(Productivity_rival) is significant at p-value<0.01.

4.2.4 An explanation for D = C

It is worth remarking that prior studies—and now, our own as well—have arrived at the conclusion of D being equal—or very nearly so—to C. We propose that if social incentives depend on the presence and strength of "rival teams," then effort levels between teams will converge.[16] In the appendix, we show that functions of competitor-dependent social incentives generate D = C over time. We express such a function as $S^m(e_i, e_j) = \theta^m(max(e_i, e_j) - e_i)$ with $0 < \theta^m < 1$ and where e_i and e_j denote the effort of the focal and competing teams, respectively. This function indicates that social incentives will be exerted only in the team that lags in effort, and it is consistent with the notion that social incentives are

[16] This does not mean that productivity levels will converge, given that these depend on many factors other than workers' effort choices, such as managerial practices, technology, managers' capacity, or input quality.

triggered by the presence of strong rivals (as discussed in the previous section). In our setting, it is not unlikely that such a functional form might emerge. Effort across teams is observable, and thus, the competitor's effort is a natural and useful benchmark.

4.2.5 B – A is necessary to unpack the mechanisms driving D – C

In this section we posit that, under general conditions, it is very difficult to disentangle the presence and/or the size of the underlying mechanisms of free-riding, effort complementarity and asymmetric social incentives for a given a result for D – C without also having an estimate for B – A in the same setting. In the appendix, we provide a mathematical proof of this claim.

The basic intuition for this result is that the estimate of B – A facilitates the assessment of free-riding in teams: the extent of free-riding is, by definition, measured by how much the impact of pay-for-performance on effort decreases as N increases. This is readily evident from the "pay-for-performance" term in the optimal effort of our model: $\frac{1+b(N-1)}{N}k_s$, where the expression $\frac{1}{N}k_s$ captures how much N reduces the impact of k_s, where k_s is roughly equal to B – A (except when monitoring is present). Thus, without B – A, it is simply not possible to obtain an accurate measure of free-riding. The same applies for effort complementarity: its size is measured by the expression $\frac{b(N-1)}{N}k_s$, and thus, if one has an estimate of b, then one also requires a proxy of k_s by measuring B – A. Finally, with the previous elements in place (i.e., B – A, size of free-riding and size of effort complementarity), if one has an estimate of D – C, one can obtain, by difference, an estimate of "asymmetry in social incentives."

To further substantiate our claim, we provide an example. Imagine that in the context of Friebel et al. (2017) there were some bakeries that could be operated by single individuals and that B – A is estimated at X% increase in performance. Whether X is equal to 20% or 3% is a significant detail. If X were 20%, then the 3% estimate for D – C obtained by Friebel et al. means that either i) free-riding *is large* and its counteracting forces—e.g., effort complementarity, asymmetric social incentives—are *small* (and thus, D – C = 3% is obtained via a low D), or that ii) both asymmetric social incentives and effort complementarity are *large*, the former being approximately 17 percentage points larger under weak compared to strong incentives (and thus, D – C = 3% is obtained via high C), or iii) a combination of these two extremes.[17] Consider now that B – A were 3%. The extent of free-riding would be much smaller, and so the rest of the forces that yield D – C = 3%. More importantly, if effort complementarity were strong (for example, reducing the impact of incentives from 3% in individuals to 2.5% in teams), then, in contrast to the case of B – A = 20%, social incentives would be bound to be either symmetric, or else asymmetric but very small. To reiterate, it is necessary to know the size of B – A in order to say anything about social incentives as a whole.

5 Discussion and conclusion

In this paper, we analyzed the impact of incentives on the respective productivity of individuals and teams based on data gathered from two competing beverages firms selling in 175 supermarkets in Santiago de Chile during two discrete periods: January to August 2011, and January 2014 to June 2015. The unique characteristics of our setting allowed us to unpack the generic mechanisms that cause the effectiveness of incentives on teams: free-riding, effort complementarity, and social incentives. Although we cannot definitively rule out endogeneity in our findings, several techniques and analyses increase confidence in our estimates. Our three main results contribute to the literature on formal incentives in teams.

First, we show that there is no statistically distinguishable difference in the productivity of incentivized and non-incentivized teams. Following Friebel et al. (2017), who observed that as far as team incentives are concerned, "the jury […] is still out," our study adds additional evidence consistent with previous findings of a null or small effect.

Second, we find that, in the same setting where incentives do not impact team productivity, incentives do work for individuals, whose productivity increases by nearly 20% (which, again, aligns with

[17] In fact, Friebel et al. (2017) report something that resonates with the logic by which non-incentivized teams also perform quite well: "senior managers were afraid that bonus payments could prove to be a financial burden on the firm. In particular, the bonus would need to be paid even to *those shop teams that would have reached their sales targets in any case*" (p. 2175; emphasis added).

the extant literature; see Ashraf and Bandiera, 2018; and Lazear, 2018). To our knowledge, ours is the first study to estimate the impact of incentives on teams and individuals in the same setting. Our novel contribution (i.e., having an estimate of B – A) is important for two reasons. First, one cannot rule out that the small or null impact of incentives documented by Friebel et al. (2017) and Delfgaauw et al. (2020 and 2022) for teams may be because formal incentives do not work at all in the particular settings studied by these authors, not even for individuals. Potential explanations for this may include culture, pervasive standardization of procedures, or informal incentives such as relational contracts. To rule out this alternative explanation, it is necessary to estimate B – A. Second, without an estimate of B – A, it is very difficult to say anything definitive about the mechanisms underlying a given result in D – C, because B – A provides a measure against which to estimate free-riding and effort complementarity.

Third and finally, we can decompose the specific mechanisms that drive our results. In our setting, the null advantage of teams with formal incentives is not because formal incentives do not work, but rather because teams do not need them: in teams without formal incentives, social incentives are much stronger, boosting such teams' productivity to the level of teams with formal incentives. The contribution here is that, by considering B – A, we are—to our knowledge—the first paper to fully and precisely unpack the mechanisms involved in the estimate of D vs. C. Guided by our model, strengthened by the richness of our setting, we show that i) pay-for-performance works because effort complementarity is large, effectively counteracting free-riding; and ii) social incentives are asymmetric, being much more impactful under weak rather than strong formal incentives. Therefore, D = C because C increases, not because D decreases.

Our findings have two important managerial implications. First, it is not obvious that firms would wish to introduce formal incentives when production is executed in teams, even if they have been effective for individuals in the same setting. Effort may increase due to such incentives, but social incentives are likely to be crowded out, generating no net advantage against non-incentivized teams; given that incentives are expensive, this would tilt the balance against their adoption. Second, to precisely gauge the suitability of introducing incentives in teams, managers must carefully assess the mechanisms of free-riding, monitoring, effort complementarity and social incentives in their particular workplace, and would tend to benefit by adopting team incentives when these mechanisms are likely to yield D > C.

Our paper is not without limitations. First, we were able to diminish—but not rule out—endogeneity concerns. Second, we do not directly test for the presence of social incentives; instead, we measured them as a residual, based on the model we built. (Nonetheless, our analysis of the RtC provides more direct evidence). The plausibility of this exercise rests on the assumption that our model is "complete" (cf. total factor productivity estimation) and as such, that there are only two elements (which are mutually exclusive and complementary exhaustible) that matter when assessing the response to formal incentives in teams: 1) the direct (selfish) effort response to formal incentives (where free-riding and effort complementarity are embedded), and 2) social incentives. There may be other elements to include in 1) and 2) that we have not considered—or perhaps an entirely new third category should be added (e.g., non-social psychological traits such as loss aversion).

6 References

Akerlof, G. A., & Kranton, R. E. (2005). Identity and the economics of organizations. *Journal of Economic Perspectives*, *19*(1), 9–32.

Alchian, A. A., & Demsetz, H. (1972). Production, information costs, and economic organization. *The American Economic Review*, *62*(5), 777–795.

Ashraf, N., & Bandiera, O. (2018). Social incentives in organizations. *Annual Review of Economics*, 10, 439–463.

Babcock, P., Bedard, K., Charness, G., Hartman, J., & Royer, H. (2015). Letting down the team? Social effects of team incentives. *Journal of the European Economic Association*, *13*(5), 841–870.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, *11*(5), 1079–1114.

Bauer, M., Blattman, C., Chytilová, J., Henrich, J., Miguel, E., & Mitts, T. (2016). Can war foster cooperation? *Journal of Economic Perspectives*, *30*(3), 249–74.

Bloom, N., & Van Reenen, J. (2011). Human resource management and productivity. In *Handbook of Labor Economics* (Vol. 4, pp. 1697–1767). Elsevier.

Böhm, R., and Rockenbach, B. (2013). The inter-group comparison-intra-group cooperation hypothesis: Comparisons between groups increase efficiency in public goods provision. *PloS One*, *8*(2)e56152.

Boning, B., Ichniowski, C., & Shaw, K. (2007). Opportunity counts: Teams and the effectiveness of production incentives. *Journal of Labor Economics*, *25*(4), 613-650.

Bowles, S. (2006). Group competition, reproductive leveling, and the evolution of human altruism. *Science*, *314*(5805), 1569–1572.

Bowles, S., & Polania-Reyes, S. (2012). Economic incentives and social preferences: Substitutes or complements? *Journal of Economic Literature*, *50*(2), 368–425.

Brahm, F., & Poblete, J. (2018). Incentives and ratcheting in a multiproduct firm: A field experiment. *Management Science*, *64*(10), 4552–4571.

Brickley, J. A., & Dark, F. H. (1987). The choice of organizational form: The case of franchising. *Journal of Financial Economics*, *18*(2), 401–420.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., & Wiesen, D. (2017). The effects of introducing mixed payment systems for physicians: Experimental evidence. *Health Economics*, *26*(2), 243–262.

Burgess, S., Propper, C., Ratto, M., Kessler Scholder, S. V. H., & Tominey, E. (2010). Smarter task assignment or greater effort: The impact of incentives on team performance. *The Economic Journal*, *120*(547), 968–989.

Chan, T. Y., Li, J., & Pierce, L. (2014). Compensation and peer effects in competing sales teams. *Management Science*, *60*(8), 1965–1984.

Corsten, D., & Gruen, T. (2003). Desperately seeking shelf availability: An examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management, 31(12: 605-617*.

Delfgaauw, J., Dur, R., & Souverijn, M. (2020). Team incentives, task assignment, and performance: A field experiment. *The Leadership Quarterly*, 31(3), 101241.

Delfgaauw, J., Dur, R., Onemu, O. & Sol, J. (2022). Team incentives, social cohesion, and performance: A natural field experiment. *Management Science*, *68*(1), 230–256.

Dickinson, D., & Villeval, M. C. (2008). Does monitoring decrease work effort?: The complementarity between agency and crowding-out theories. *Games and Economic Behavior*, *63*(1), 56–76.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–78.

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation? *Management Science*, *59*(7), 1496–1513.

Erev, I., Bornstein, G., & Galili, R. (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology*, *29*(6), 463–478.

Francois, P., Fujiwara, T., & Van Ypersele, T. (2018). The origins of human prosociality: Cultural group selection in the workplace and the laboratory. *Science Advances*, 4(9), eaat2201.

Frank, D. H., & Obloj, T. (2014). Firm-specific human capital, organizational incentives, and agency costs: Evidence from retail banking. *Strategic Management Journal*, *35*(9), *1279*–1301.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, *87*(4), 746–755.

Friebel, G., Heinz, M., Krueger, M., & Zubanov, N. (2017). Team incentives and performance: Evidence from a retail chain. *American Economic Review*, *107*(8), 2168–2203.

Garton, E., & Noble, A. (2017). How to make agile work for the C-suite. *Harvard Business Review*, 19, 2017.

Gaynor, M., & Gertler, P. (1995). Moral hazard and risk spreading in partnerships. *The RAND Journal of Economics*, 591–613.

Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., ... & Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, *332*(6033), 1100–1104.

Gibbons, R. (2005). Incentives between firms (and within). *Management Science*, 51(1), 2–17.

Gibbons, R., & Roberts, J. (2012). Economic theories of incentives in organizations. In *The Handbook of Organizational Economics* (pp. 56–99). Princeton University Press.

Grosse, S., Putterman, L., & Rockenbach, B. (2011). Monitoring in teams: Using laboratory experiments to study a theory of the firm. *Journal of the European Economic Association*, *9*(4), 785–816.

Gubler, T., Larkin, I., & Pierce, L. (2016). Motivational spillovers from awards: Crowding out in a multitasking environment. *Organization Science*, *27*(2), 286–303.
Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, *111*(3), 465–497.
Hansen, D. G. (1997). Worker performance and group incentives: A case study. *ILR Review*, *51*(1), 37–49.
Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics*, *30*(4), 637–646.
Herbst, D., & Mas, A. (2015). Peer effects on worker output in the laboratory generalize to the field. *Science*, *350*(6260), 545–549.
Holmström, B. (1982). Moral hazard in teams. *The Bell Journal of Economics*, 324–340.
Itoh, H. (1991). Incentives to help in multi-agent situations. *Econometrica*, 611–636.
Jensen, N., Lyons, E., Chebelyon, E., Le Bras, R., & Gomes, C. (2020). Conspicuous monitoring and remote work. *Journal of Economic Behavior & Organization*, *176*, 489–511.
Kandel, E., & Lazear, E. P. (1992). Peer pressure and partnerships. *Journal of Political Economy*, *100*(4), 801–817.
Kaplan, S., & Henderson, R. (2005). Inertia and incentives: Bridging organizational economics and organizational theory. *Organization Science*, *16*(5), 509–521.
Kozlowski, S. W. J., & Bell, B. S. (2013). Work groups and teams in organizations. In N. W. Schmitt, S. Highhouse, & I. B. Weiner (Eds.), *Handbook of Psychology: Industrial and Organizational Psychology* (pp. 412–469). John Wiley & Sons, Inc.
Kretschmer, T., & Puranam, P. (2008). Integration through incentives within differentiated organizations. *Organization Science*, *19*(6), 860–875.
Larkin, I., Pierce, L., & Gino, F. (2012). The psychological costs of pay-for-performance: Implications for the strategic compensation of employees. *Strategic Management Journal*, *33*(10), 1194–1214.
Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, *90*(5), 1346–1361.
Lazear, E. P., & Oyer, P. (2012). Personnel economics. In *The Handbook of Organizational Economics* (pp. 479–519). Princeton University Press.
Lazear, E. P. (2018). Compensation and incentives in the workplace. *Journal of Economic Perspectives*, *32*(3), 195–214.
Lee, S., & Meyer-Doyle, P. (2017). How performance incentives shape individual exploration and exploitation: Evidence from microdata. *Organization Science*, *28*(1), 19–38.
Lee, S., & Puranam, P. (2017). Incentive redesign and collaboration in organizations: Evidence from a natural experiment. *Strategic Management Journal*, *38*(12), 2333–2352.
Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.
Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112–45.
Mitsuhashi, H., & Nakamura, A. (2022). Pay and networks in organizations: Incentive redesign as a driver of network change. *Strategic Management Journal*, articles in advance.
Nagin, D. S., Rebitzer, J. B., Sanders, S., & Taylor, L. J. (2002). Monitoring, motivation, and management: The determinants of opportunistic behavior in a field experiment. *American Economic Review*, *92*(4), 850–873.
Nalbantian, H. R., & Schotter, A. (1997). Productivity under group incentives: An experimental study. *American Economic Review*, 314–341.
Nyberg, A. J., Maltarich, M. A., Abdulsalam, D. D., Essman, S. M., & Cragun, O. (2018). Collective pay for performance: A cross-disciplinary review and meta-analysis. *Journal of Management*, *44*(6), 2433–2472.

Obloj, T., & Zenger, T. (2017). Organization design, proximity, and productivity responses to upward social comparison. *Organization Science*, *28*(1), 1–18.
Obloj, T., & Zenger, T. (2019). (Unavoidable) dynamics in incentive design. In *Handbook of Research on Strategic Human Capital Resources*. Edward Elgar Publishing.
Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.
Pierce, L., Snow, D. C., & McAfee, A. (2015). Cleaning house: The impact of information technology monitoring on employee theft and productivity. *Management Science*, *61*(10), 2299–2319.
Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature, 37*(1), 7–63.
Presbitero, A. F., & Rabellotti, R. (2014). Geographical distance and moral hazard in microcredit: Evidence from Colombia. *Journal of International Development*, *26*(1), 91–108.
Puranam, P. (2018). *The Microstructure of Organizations*. Oxford University Press.
Reiner, G., Teller, C., & Kotzab, H. (2013). Analyzing the efficient execution of in-store logistics processes in grocery retailing—The case of dairy products. *Production and Operations Management, 22*(4), 924–939.
Schweitzer, M. E., Ho, T. H., & Zhang, X. (2018). How monitoring influences trust: A tale of two faces. *Management Science*, *64*(1), 253–270.
Siemsen, E., Balasubramanian, S., & Roth, A. V. (2007). Incentives that induce task-related effort, helping, and knowledge sharing in workgroups. *Management Science*, *53*(10), 1533–1550.
Soetevent, A. R. (2006). Empirics of the identification of social interactions: An evaluation of the approaches and their results. *Journal of Economic Surveys*, *20*(2), 193–228.
Ton, Z., & Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*, *19*(5), 546–560.
Weiss, A. (1987). Incentives and worker behavior: Some evidence. *NBER working paper,* (w2194).
Wooldridge, J. M. (2002). Econometric analysis of cross section and panel data. MIT press. *Cambridge, MA*, *108*.

Online appendix

Mathematical proofs

Proof of Prediction 1. When $N = 1$, we have only individual production. Using ${e_i}^*$, we find that computing B – A yields $k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}\right) > 0$ ■

Proof of Prediction 2.

Part 1. When $N > 1$, we have team production. If social incentives are symmetric, using ${e_i}^*$we find that D – C is equal to $\frac{1+b(N-1)}{N} k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}\right)$ and that (B –A) – (D – C) is equal to $\left(1 - \frac{1+b(N-1)}{N}\right) k_s$. If b < 1, D – C difference becomes smaller as N increases; however, it is not equal to zero, given that $\left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}\right) > 0$. If b < 1, then (B –A) – (D – C) > 0 ■

Part 2. When $N > 1$, we have team production. If social incentives are symmetric, using ${e_i}^*$, we find that D – C yields $\frac{1+b(N-1)}{N} k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}\right) + \left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}\right)$ and that (B –A) – (D – C) is equal to $\left(1 - \frac{1+b(N-1)}{N}\right) k_s + \left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i}\right)$. In order for D – C to be equal to or smaller than zero (and thus, C ≥ D), there is a threshold value for $S^{strong}(e_i, e_{-i}, N) < S^{weak}(e_i, e_{-i}, N)$ such that this condition holds: $\frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} \geq \frac{1+b(N-1)}{N} k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}\right)$; the higher the value of b, the higher this threshold value needs to be. Given Prediction 1, this threshold value implies that (B –A) – (D – C) ≥ 0 ■

Proof of Prediction 3. Using ${e_i}^*$, we find that computing A – C yields $\frac{\partial w^{weak}(e_i)}{\partial e_i} - \left(\frac{\partial w^{weak}(e_i)}{\partial e_i} + \frac{\partial S^{weak}(e_i,e_{-i})}{\partial e_i}\right) = -\frac{\partial S^{weak}(e_i,e_{-i})}{\partial e_i} < 0$ ■

Proof of Prediction 4. Using ${e_i}^*$, we verify that computing D – B yields $\left(\frac{1+b(N-1)}{N} k_s + \frac{\partial w^{strong}(e_i)}{\partial e_i} + \frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}\right) - \left(k_s + \frac{\partial w^{strong}(e_i)}{\partial e_i}\right) = k_s\left(\frac{1+b(N-1)}{N} - 1\right) + \frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}$. The first term is negative when $N > 1$ and $b < 1$; thus, if $\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i}$ is large enough, the full expression can be positive ■

<u>Proof of the impossibility of making definitive claims about mechanisms of D – C without an estimate of B – A.</u> Let us assume that social incentives are an aggregate of a very large set of specific mechanisms s_i^m at play (e.g., peer pressure, reciprocal altruism, inequity aversion, envy, status-seeking, altruism, trust, relational contracting, reputation, social cohesion, prosocial leadership) and that it is not feasible to directly measure each of them. That is, $S^m(e_i, e_{-i}, N) = \sum_i^I s_i^m(e_i, e_{-i}, N)$, where I is large and there exists i such that s_i^m is not measureable. Using $e_i{}^*$, we obtain

$$D - C = \frac{1+b(N-1)}{N} k_s + \left(\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i} \right) + \left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i} \right) \quad (1)$$

Assume that $\frac{\partial w^{strong}(e_i)}{\partial e_i} - \frac{\partial w^{weak}(e_i)}{\partial e_i}$ is an observable constant $M > 0$ and that it would be possible to estimate effort complementarity b.

If the setting allows us to estimate k_s as:

$B - A = k_s + M \Rightarrow k_s = B - A - M$ (2)

Then, by replacing (2) in (1), we can estimate the impact of asymmetric social incentives.

However, if we assume that $B - A$ is unknown, then two unknowns remain in the formulation of D – C, namely the impact of pay-for-performance k_s and the asymmetric social incentives $\left(\frac{\partial S^{strong}(e_i,e_{-i},N)}{\partial e_i} - \frac{\partial S^{weak}(e_i,e_{-i},N)}{\partial e_i} \right)$. Thus, there are an infinite number of combinations of these two variables that satisfy (1) and it is impossible to estimate which mechanism drives the $D - C$ estimation. Solving this problem by attempting to measure “social incentives” and then finding k_s is not feasible: given that the number of specific mechanisms s_i^m is large, and that at least one is not measurable, we still obtain infinite solutions.

■

Proof of effort convergence when social incentives are competitor-dependent. Let us assume that team i is given strong incentives and team j, weak incentives. Both teams will then exert the following optimal efforts (without any social incentives, as efforts are unobservable at first):

$$e_i = \frac{1+b(N-1)}{N} k_s + \frac{\partial w^{strong}(e_i)}{\partial e_i} \quad ; \quad e_j = \frac{\partial w^{weak}(e_j)}{\partial e_j}$$

Once efforts are observed in the two teams, social incentives can be felt by the team with the lowest effort. We introduce a social incentives function of the form $S^m(e_i, e_j) = \theta^m(max(e_i, e_j) - e_i)$, which is continuously felt and updated by the team with the lowest effort. We assume that this updating process occurs over n periods of time as long as there is a difference in efforts.

In Period 1, given that $e_i > e_j$, only team j will feel social incentives equal to $S^{weak}(1) = \theta^{weak}(e_i - e_j) = \theta^{weak}\ \varepsilon(0)$, with $\varepsilon(0)$ being the initial difference in efforts. As $0 < \theta^{weak} < 1$, this difference of efforts could not be fully offset by the social incentives felt in Period 1, and we can write the remaining difference in efforts as $\varepsilon(1) = (1 - \theta^{weak})\varepsilon(0)$.

In Period 2, only team j will experience social incentives $S^{weak}(2) = \theta^{weak}\varepsilon(1) = \theta^{weak}(1 - \theta^{weak})\varepsilon(0)$. With a similar reasoning as above, we can write the difference in optimal efforts in Period 2 that was not offset by the social incentives as $\varepsilon(2) = (1 - \theta^{weak})\varepsilon(1) = (1 - \theta^{weak})^2\varepsilon(0)$.

As n grows, in Period n, team j will experience social incentives equal to $S^{weak}(n) = \theta^{weak}\varepsilon(n-1) = \theta^{weak}(1 - \theta^{weak})^{n-1}\varepsilon(0)$.

The total social incentives felt by team j over n periods of time is equal to:

$$\sum_{i=1}^{n} S^{weak}(i) = S^{weak}(1) + S^{weak}(2) + \cdots + S^{weak}(n)$$

$$= \theta^{weak}\ \varepsilon(0) + \theta^{weak}(1 - \theta^{weak})\varepsilon(0) + \cdots + \theta^{weak}(1 - \theta^{weak})^{n-1}\varepsilon(0)$$

$$= \theta^{weak}\ \varepsilon(0)(1 + (1 - \theta^{weak}) + \cdots + (1 - \theta^{weak})^{n-1})$$

As $0 < \theta^{weak} < 1$, and $n \to \infty$, here we recognize a Taylor expansion series such that:

$$\lim_{n\to\infty}\sum_{i=1}^{n} S^{weak}(i) = \theta^{weak}\ \varepsilon(0)\left(\frac{1}{1-(1-\theta^{weak})}\right) = \varepsilon(0) = e_i^* - e_j^* \blacksquare$$

Figures

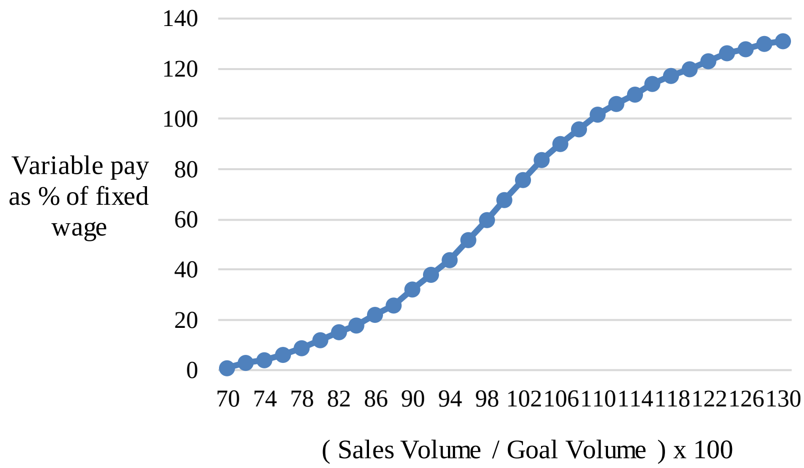

Figure A1. Example of a payout curve

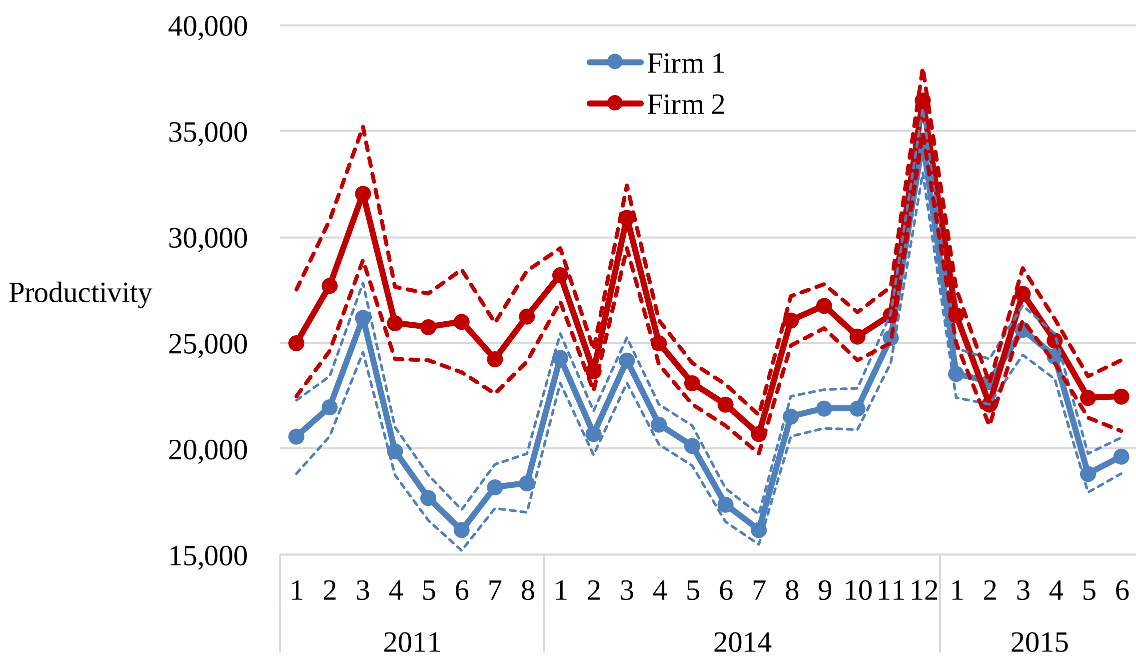

Figure A2. Productivity over time (The dotted lines display a 95% confidence interval)

Tables

Table A1. Adding time-variant controls.

	Dependent variable: Ln (Productivity)			
	(1)	(2)	(3)	(4)
Weak incentives	-0.673 (0.234)	-0.888 (0.224)	-0.447 (0.227)	-0.448 (0.259)
Weak incentives x Store size	0.086 (0.026)	0.075 (0.028)	0.066 (0.026)	0.064 (0.027)
Distance to store	-0.052 (0.062)			-0.042 (0.065)
Weak incentives x Distance to store	0.033 (0.042)			0.032 (0.039)
HHI product categories		0.216 (0.131)		0.281 (0.133)
Weak incentives x HHI product categories		0.495 (0.399)		0.321 (0.370)
% “returnable” sales			-0.418 (0.307)	-0.489 (0.307)
Weak incentives x % “returnable” sales			-2.911 (0.664)	-2.950 (0.661)
Period FE?	YES	YES	YES	YES
Firm x Store FE?	YES	YES	YES	YES
Constant?	YES	YES	YES	YES
R-square	53.91%	55.6%	56.3%	55.9%
Observations	8,812	8,817	8,789	8,593
Robust standard errors, clustered at the firm–store level.				

Table A2. Adding store-level controls.

	Dependent variable: Ln (Productivity)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Weak incentives	-0.689 (0.226)	-0.605 (0.280)	-0.811 (0.217)	-0.892 (0.227)	-0.968 (0.220)	-0.700 (0.234)	-0.816 (0.214)	-0.868 (0.282)
Weak incentives x Store size	0.063 (0.033)	0.077 (0.028)	0.081 (0.027)	0.085 (0.027)	0.088 (0.026)	0.083 (0.028)	0.074 (0.027)	0.076 (0.032)
Weak incentives x Warehouse size	0.033 (0.032)							0.019 (0.037)
Weak incentives x Area manager		-0.031 (0.028)						-0.007 (0.030)
Weak incentives x Procedures			0.028 (0.027)					0.014 (0.034)
Weak incentives x Store conditions				0.038 (0.022)				0.029 (0.034)
Weak incentives x Warehouse conditions					0.046 (0.025)			0.042 (0.037)
Weak incentives x Work-tools						-0.015 (0.026)		-0.058 (0.032)
Weak incentives x Competition level							0.042 (0.027)	0.029 (0.029)
Period FE?	YES	YES	YES	YES	YES	YES	YES	YES
Firm x Store FE?	YES	YES	YES	YES	YES	YES	YES	YES
Constant?	YES	YES	YES	YES	YES	YES	YES	YES
R-square	54.4%	54.5%	54.5%	54.5%	54.5%	54.5%	54.5%	54.8%
Observations	8,962	8,962	8,962	8,962	8,962	8,962	8,962	9,962

Robust standard errors in parentheses, clustered at the firm-store level. The variable “warehouse size” captures the natural logarithm of the square meters of the store’s warehousing/stock-piling area devoted to liquids (this area is not accessible to customers). The remaining variables capture the degree of agreement of the Firm 2 store supervisor in mid-2011, using a five-point Likert scale based on supervisors’ replies to six prompts. For *Area manager,* the prompt was: “The manager of the beverages area and its boss have been helpful for the proper functioning of replenishment”; for *Procedures*: “The rules, procedures and schedules of the store benefit the proper functioning of replenishment”; for *Store conditions*: “The physical conditions of the store are not ideal: long distances, reduced spaces in hallways, bad lighting, etc.”; for *Warehouse conditions*: “The physical conditions of the store’s warehouse are not ideal: not enough room for stockpiling, reception of orders, and inventory management”; for *Work-tools*: “The tools that are used in the replenishment tasks (carts, forklifts, etc.) are of low quality, insufficient and must be constantly disputed with other suppliers of the store”; and for *Competition level,* the prompt was: “The level of competition in replenishment is high; competitors are aggressive, persistent, and very clever to fight and win in terms of product visibility and additional spaces.”

Table A3. Robustness using the overlapping agency.

	Dependent variable: Ln (Productivity)					
	(1)	(2)	(3)	(4)	(5)	(4)
Weak incentives	-0.126 (0.048)	-0.153 (0.055)	-0.715 (0.021)	-0.703 (0.021)	-0.746 (0.215)	-0.710 (0.259)
Weak incentives x Store size			0.077 (0.027)	0.075 (0.027)	0.077 (0.027)	0.072 (0.032)
Common agency	0.104 (0061)	0.086 (0.061)	0.092 (0.062)	-0.938 (0.400)	-0.898 (0.391)	-0.877 (0.390)
Common agency x Store size				0.134 (0.050)	0.127 (0.048)	0.124 (0.049)
Common agency x Weak incentives		0.074 (0.055)			0.069 (0.053)	-0.059 (0.043)
Common agency x Weak incentives x Store size						0.016 (0.054)
Period FE?	YES	YES	YES	YES	YES	YES
Firm x Store FE?	YES	YES	YES	YES	YES	YES
Constant?	YES	YES	YES	YES	YES	YES
R-square	54.41%	54.47%	54.66%	54.88%	54.93%	
Observations	9,008	9,008	9,008	9,008	9,008	9,008
Robust standard errors in parentheses, clustered at the firm-store level.						

Table A4. Worker characteristics across store sizes for Firm 2 in year 2011.

Variable	Mean, (standard deviation), and [observations] in small stores, defined as stores below median size	Mean, (standard deviation), and [observations] in small stores, defined as stores above median size	p-value (null: no difference in means)
Age	30.61 (9.5) [301]	30.94 (4.81) [337]	0.58
Tenure (in days)	627 (437) [393]	610 (276) [404]	0.52
Percentage of married workers	34.31% (39.5%) [307]	31.29% (34.39%) [339]	0.26
Percentage of wage from overtime	2.60% (2.13%) [405]	2.42% (2.19%) [416]	0.23
Percentage of wage paid in advance	14.90% (4.31%) [405]	14.62% (4.17%) [416]	0.34