ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
CAMPUS DI FORLÌ

AICCON

Working Paper 193

No recognition without participation: a field experiment in the workplace

Luca Savorelli, *University of St Andrews*
Giulio Ecchia, *University of Bologna*
Raimondello Orsini, *University of Bologna*

+39 0543 62327
ecofo.aiccon@unibo.it
www.aiccon.it

aiccon
cooperazione | non profit

AICCON – Italian Association for the Promotion of the Culture of Co-operation and of Nonprofit is an association formed in 1997 among the **University of Bologna, Faculty of Economics, Forlì Campus**, within the academic course on Social Economy. The aim of the Association is to encourage, support, and organise initiatives to promote the culture of solidarity with particular attention to idealities, perspectives, activities, and problems connected to Nonprofit Organizations and Co-operative Enterprises.

AICCON is part of network of people and institutions (**EMES Network**) at national and international level that, starting from its members, forms the environment in which it is located. AICCON, throughout the years, has increased its reach and succeeds to the local, national and international context in which it works.

No recognition without participation: a field experiment in the workplace.*

Luca Savorelli[a,1], Giulio Ecchia[b,2], Raimondello Orsini[b,3]

[a] *University of St Andrews*
[b] *University of Bologna*

January 28, 2022

Abstract

Firms are increasingly pursuing employees' participation programmes, yet field evidence on their impact on performance is scant. To study their effect on quality provision, we conduct a field experiment that exogenously determines whether workers participate or not in the design of a recognition scheme. With no awareness of being part of a research study, some workers vote on the scheme's format and ranking system, whereas others have no voice. The results show that the scheme backfires without employees' participation: mistakes increase by around 50% in comparison with a control group and with employees who participated in the design of the scheme. The experiment also exogenously determines the timing of the onset and of the withdrawal of the recognition scheme, showing that adverse outcomes persist even after the scheme's end. These adverse effects are driven by mistakes affecting the organization's management, rather than end-users or colleagues. Employees' performance responds directly to experimental manipulation and Hawthorne-type effects operate separately from the participation mechanism.

Keywords: natural field experiment, participation, recognition, incentives, persistence, Hawthorne effect.

JEL: D23, J32, I19, M52

*Luca Savorelli gratefully acknowledges financial support from the Scottish Institute for Research in Economics. Thanks to Federico Fornasari and Kelly McCoy for excellent research assistantship. We thank Stefano DellaVigna, Imran Rasul, David Ulph, and participants to the Royal Economic Society Annual Conference 2019; Arne Ryde Workshop on Experimental Methods in the Study of Firms, Management and Entrepreneurs 2017; London Experimental Workshop 2016; International Field Days Workshop 2015; to seminars at the Universities of St Andrews and Bologna for useful comments. Ethical approval obtained by the University Teaching and Research Ethics Committee (UTREC) at the University of St Andrews. The usual disclaimer applies.

[1] *Corresponding author*, School of Economics and Finance, The Scores, KY16 9AR, Scotland, United Kingdom. luca.savorelli@st-andrews.ac.uk

[2] Department of Economics, Piazza Scaravilli 1, 40126 Bologna, Italy. giulio.ecchia@unibo.it

[3] Department of Economics, Piazza Scaravilli 1, 40126 Bologna, Italy. raimondello.orsini@unibo.it

1. Introduction

Firms are increasingly pursuing participation programmes which broaden the extent of employees' consultation or their autonomy in decision-making, with some studies reporting that 40% of employees in the US and 34% in the UK have influence on firm's decisions and their job (Park et al., 2010; Van Wanrooy et al., 2013). On the one hand, through these programmes a firm may gain informational advantage by shifting decision power to the employees, who have direct experience with the job's tasks and problems (Argarwal et al., 2018). On the other hand, through a motivational effect, participation may increase employees' satisfaction, reduce their cost of effort and improve their performance (Cassar and Meier, 2018). The Covid-19 pandemic has forced a vast majority of organisations to increase the autonomy of workers, who were often required to work from home, and it heated the debate on workers' participation in the workplace decision-making. A greater involvement of workers in the post-pandemic world has since been advocated by international organizations (the World Health Organization, the International Labour Organization, the European Agency for Safety and Health, see Cockburn 2020). In the private sector some, such as McKinsey, have called to reject hierarchical models in favour of greater engagement with stakeholders (Alexander et al., 2020).

Despite the quickly growing interest in this topic, evidence from the field on the effects of increased participation on workers' performance is hard to find.[4] Two exceptions are the pioneering framed field experiments by Lawler and Hackman (1969) and Scheflen et al. (1971), which show a positive effect of participation on work attendance.[5] A consolidated literature has studied monetary, intrinsic, and social incentives for employees through field experiments, but such evidence focuses only on cases where the managers introduce incentives top-down, without workers' consultation.[6] This paper attempts to narrow this gap in

[4] A recent growing literature studies a related, but distinct, topic, i.e. "codetermination" agreements, where elected workers representatives are included both in firms' boards and in shop-floor level committees (for a review, see Jäger et al. 2021), but experimental or quasi-experimental evidence about the impact on performance of strong forms of shop-floor codetermination is also scant (Jäger et al., 2021).

[5] Lawler and Hackman (1969) found that the percent of scheduled worked hours increased only in the participation treatment group. The study, however, is based on descriptive statistics and does not include any econometric analysis. The workers (cleaners) knew that they were part of an experiment and the authors were directly involved in focus groups. A follow-up experiment in Scheflen et al. (1971) shows that analogous positive effects persisted also in the long run in the participation treatment group, but not for all workers.

[6] Intrinsic incentives are grounded on the individual's desire to perform a task for its own sake. Social incentives are driven by the relationships with co-workers in the workplace. A recent economic literature has shown that social incentives might reduce free-riding and favour socially connected workers, while performance-based incentives favour high-ability workers (e.g., Bandiera et al. 2009, 2013). In some of these settings, the net effect of social incentives can be detrimental to a firm's performance.

the literature and its results show that the success of incentives is influenced by the way in which they are introduced.

We analyse the effect of participation on performance by conducting a natural field experiment in the workplace, where the firms' management agreed with us to introduce a novel recognition scheme. The experiment lasted for fourteen weeks, involved around 110 workers employed in three care homes, collecting around 1 500 observations over about 44 000 working hours. The three facilities were all part of a single company that operates multiple care homes and were chosen to be comparable in terms of size, infrastructures, logistics, composition of carers and people being cared for. Until being debriefed after the end of the experiment, the workers did not know that they were part of an academic study. They thus framed the introduction of the recognition scheme as part of matter-of-course operations.

The experimental design's main feature is to exogenously determine the timing of the onset *and* of the withdrawal of a recognition scheme, varying the process through which it is introduced: in the first care home in a participative, bottom-up form; in the second care home, in a top-down, hierarchical form. In the third care home, which is our control group, no recognition scheme or other incentive is introduced. In the bottom-up care home, the area manager invited the employees to a meeting, where she presented the introduction of the recognition scheme and offered the employees to vote on supposedly secondary features (the "Bottom-up" treatment): its communication format (public or private) and the rule to determine who receives recognition. Through collective choice, the employees' voted for a private meeting the principal care-home manager should congratulate the employees who provide the highest quality of performance (measured by the lowest number of weekly mistakes in the job). In the top-down care home, the same area manager performed the same presentation of the introduction of the same recognition scheme, but without letting the workers vote on any feature (the "Top-down" treatment).[7]

This experiment is not a randomised control trial, as a meaningful randomisation would have implied that, in the same care home, some workers were in the treatment group and

[7]The employees in the "Top-down" care home did not know that some features of the scheme were chosen by the workers in the other care home, and only then also implemented there. The geographical distance between the care homes ensured that there was no communication between the workers in the two care homes. Also, we carefully instructed the area manager that ran the two meetings to deliver them in the same format. We also made sure that, except for the choice over some features in the "Bottom-up" care home, the information about the scheme was the same in the two meetings, so that results may not be driven by better information about the scheme obtained during the meeting in either care home. Employees did not have voice on the introduction of the scheme itself in neither care home, and did not raise concerns about this.

others in the control group, biasing the treatment effect. The experimental design avoids the issue of contamination and allows us to study the results from the angle of both a within-worker analysis *and* a difference-in-differences analysis. To rule out spurious effects which may be driven by unobservable factors occurring at the same time of the treatment, we exploit the fact that we exogenously control both the introduction and the withdrawal of the scheme, we secretly observe the workers performance before any treatment is introduced, and we provide other evidence, such as the analysis of changes in the *types* of mistakes during the treatments' phase.

Quality is the most important and variable dimension of performance of the healthcare assistants' job. It was measured through a careful record of the number of mistakes and very granular details of their nature.[8] In the weeks before the core treatments, we obtained a measurement of the healthcare assistants' baseline performance by secretly recording their mistakes. After the managers had disclosed the introduction of the new form to record mistakes, we controlled for "Hawthorne-type" effects, i.e. the possibility that the change in performance is driven chiefly by the greater attention to mistakes, even if work conditions and incentives do not change (Levitt and List, 2011).[9] In addition, after discontinuing the recognition scheme in both care homes, we continued recording for five weeks the employees' performance to examine the persistence of the treatments' effects.

A key feature of this context is that broadening the employees' decision-making does not confer any informational advantage, which makes it an ideal setting to neatly study employees' motivation. Also, as the quantitative component of the job's performance (the number and type of operations) is pre-determined by a strict protocol and tight timing of tasks (e.g. the number and timing of beds' preparation is fixed), this setting allows us to focus on variability along the quality dimension, which plays the most important role for the end-users' service and the management operations.[10] The context of the experiment itself is also a source of novelty of this contribution: we analyse the performance of healthcare

[8] For example, in data reporting that the healthcare assistant failed to take care of the users' appearance, it can be distinguished whether it affected the hairs, beard, nails, or other.

[9] In fact, Levitt and List (2011) show that in the original Hawthorne's experiment the evidence of Hawthorne-type effects was more nuanced than earlier thought and pertained to experimental manipulation of light, compared to natural fluctuations. Before the experiment had started, mistakes were informally reported and employees would have received a formal and informal warnings for some. However, an accurate record of minor mistakes was not kept, then. Hence, independently of their choice, after the start of the experiment workers faced the introduction of a new and supplementary format of monitoring by their supervisors (the record of the number and type of even minor mistakes), but with no incentives or penalties attached, at first.

[10] Focusing on the quality of delivery is key in healthcare services, see for example Kolstad (2013) for the effect of quality record cards on a surgeon performance.

assistants in a large Italian non-profit organization, which delivers its services in care homes for the elderly, i.e. a complex job in the service sector of an advanced economy.

We ground this paper's behavioural hypotheses in Self-Determination Theory (SDT, Deci and Ryan 2000) and its application to work motivation (Gagné and Deci 2005). According to SDT, in the context of the workplace, if workers participate in the design or choice of reward schemes, they perceive incentives as enhancing their autonomy, which leads to aligning the job's goals with their own personal goals, reducing the burden associated with completing tasks, and increasing their well-being. By contrast, if workers are not involved, these schemes may backfire and disengage the employees, as they perceive them as controlling and not aligned with their goals.[11] Within this conceptual framework, the paper's key hypotheses are that (i) performance is better in the Bottom-up treatment than in the Top-down treatment, and than in the control; (ii) the scheme backfires in the Top-down care home, worsening the employee's performance in comparison with the control care home.

This paper delivers three main results. First, using a diff-in-diffs analysis we show that, absent employee participation, the recognition scheme backfires: in the Top-down treatment the number of mistakes per worker is around 50% higher in comparison with the control group. By contrast, in the Bottom-up treatment, the recognition scheme does not backfire. In the Bottom-up treatment the number of mistakes is around 50% lower than in the Top-down treatment, but the performance is similar to the control group. We also find that mistakes decline after the withdrawal of the recognition scheme: this provides evidence that the changes in performance respond to the experimental manipulation, rather than to shocks due to unobservable factors happening at the same time of the introduction of the scheme. Even if mistakes decrease with respect to the treatment phase, after the withdrawal of the recognition scheme workers in the Top-down treatment continue to perform worse in comparison with their pre-experiment performance and with the other care homes.

Second, in comparison with both the Bottom-up care home and the control care home, the areas where the increase in mistakes is observed in the Top-down group mostly pertain to internal organization procedures, and not to services to end-users or behaviours towards colleagues. This is not mirrored in the Bottom-up care home, i.e. we do not observe a

[11] SDT's rationale is consistent also with the firms' narrative for introducing participation programmes, as presented above. SDT's line of research is among the most developed theories and provides accurate predictions, but since the 1940s (e.g., Coch and French 1948) a large body of theoretical economics, management and psychology literature investigates the drivers of employees' participation in organizations, such as: fairness concerns (Lind and Tyler, 1988; Frey et al., 2004), intrinsic values of decision rights (Bartling et al., 2014), gift-exchange mechanisms (Kube et al., 2012), procedural fairness (Thibaut and Walker, 1975) or "voice effect" (Folger, 1977).

reduction in mistakes specifically related to organizational procedures. The findings suggest that the increase in mistakes is driven by disengagement with the tasks due to the controlling effect of the scheme, and not by a form of reciprocity or retaliation against the management.

Third, we also provide evidence about Hawthorne-type effects: the sole introduction of a new system to record mistakes, with no incentives attached, drastically reduces the number of mistakes up to 50% of the baseline level in the first weeks of the experiment and is persistent.[12]

Except for the field experiments by Lawler and Hackman (1969) and Scheflen et al. (1971), to the best of our knowledge there is no study which focusses on the effect of participation on performance in a workplace in the field.[13] This paper thus provides novel findings in the area of economics field experiments, also in relation to the nature of the job and incentives. The existing literature has mainly focused on relatively simple jobs (such as fruit picking, see the series of field experiments by Bandiera et al.) and scant field evidence exists about incentives in more complex environments, which account for many jobs.

Economists have recently tested in the laboratory the role of decision rights in the utility function of agents (Dal Bó et al., 2019; Bartling et al., 2014; Fehr et al., 2013; Dal Bó et al., 2010). Two laboratory experiments study the effect of participation, which takes the form of choosing among compensation schemes. Dal Bó et al. (2019) study a real-effort task (additions), where some subject can vote for a fixed wage or a piece rate based on correct answers, and find no effect of participation on performance. Mellizo et al. (2014) study the effect of voting for a compensation scheme (tournament or revenue sharing), finding a positive effect of democratization on performance in a similar real-effort task (additions). Though increasing participation is generally expected to increase productivity, this evidence from the laboratory is mixed. We contribute to this experimental literature providing evidence from a real job in the field and focusing on non-monetary, rather than monetary, incentives.

This paper relates also to the growing literature on recognition in the workplace (e.g., Bradler et al. 2016; Ashraf et al. 2014, for a review see Ellingsen and Johannesson 2007), which delivers mixed results on the impact of recognition on performance. Following Bradler

[12] This finding is consistent with other evidence of Hawthorne-type effects in the healthcare sector (Leonard and Masatu, 2006; Leonard, 2008; Leonard and Masatu, 2010), as well as in factories (Hong et al., 2018) and household electricity consumption (Attari et al., 2015). We also find that, though the Hawthorne effect kicks in with a lag, it persists until the end of the experiment (as in Hong et al. 2018).

[13] One of the possible reasons is that the general difficulty in running field experiments, and even more so when it is difficult to empirically disentangle the performance improvement due to better information from the performance improvement due to higher workers' motivation. This is why we focus on a context where the participation scheme does not entail any informational advantage.

et al. (2016), we choose to focus on a specific form of recognition, i.e. simple congratulations to the best workers. The results show that a symbolic reward might have positive or adverse effects depending on the extent of the employees' participation in its design. This finding suggests that future academic research should investigate and devote greater attention to procedures to introduce incentives. If this is not the case, some outcomes might be incorrectly ascribed to the mechanics of the schemes themselves, rather than to the extent of workers' participation in their design. Finally, this paper also provides important pragmatic insights for the implementation and management of incentives.

The paper is structured as follows. Section 2 presents information about the organization and experimental subjects. The third section contains the theoretical foundations and the behavioural predictions of the experiment. Section 4 is dedicated to the experimental design. Sections 5 and 6 present empirical results and Section 7 concludes.

2. Data and context

In this section we describe the sources of data and the characteristics of the facilities and employees involved in the experiment. We rely on two main sources of data. The first source is personnel data of detailed socio-demographic and professional training information about each worker and unit-leader, which we will present in Section 2.1. The personnel data also include detailed information about *actual* working hours, shifts, leave and overtime work. As it will be shown in the next sections, to maximize the comparability of the care homes, we selected three of the eleven facilities of the firm on the basis of the overall balance of their characteristics: employees number, education, and professional training; number and type of residents, architectural features and interiors; staffing of nurses. All three facilities belong to the same large non-profit organization and thus share the same institutional features and contractual arrangements.

The second source is primary data on healthcare assistants' mistakes which we gathered over 14 weeks, from January 26^{th} to April 26^{th}, 2015. The period was accurately chosen to avoid seasonal effects that could arise from May due to workers' holidays and weather change. We will present details about the recording of mistakes in Section 2.4.

2.1. Balance of the employees characteristics across care homes

Table 1 reports descriptive statistics of employees' characteristics sourced from personnel data. We include in this analysis only permanent workers and we excluded from the analysis occasional workers, which are typically employed for few days. The number of workers included in the analysis is 37 for the Control and Top-down care homes and 36 in the Bottom-up care home. Key observable characteristics that could influence the number of

mistakes of employees are their professional training in the job, overall education, salary level.

Table 1: Personnel data and socio-demographic statistics

	(1) Control	(2) Bottom-up	(3) Top-down	(4) Mann-Whitney (Kruskal-Wallis)
Years of schooling	9.62	9.44	8.89	0.273
	(2.37)	(2.44)	(2.11)	(0.504)
Years of employment	8.30	9.69	9.69	0.448
	(2.79)	(2.56)	(4.17)	(0.000)
Hourly pay	8.06	8.10	8.01	0.839
	(0.37)	(0.16)	(0.37)	(0.637)
Age	45.70	40.44	45.05	0.061
	(9.00)	(10.71)	(8.98)	(0.066)
Female	0.89	0.83	1.00	
	(0.31)	(0.38)	(0.00)	
Native	0.51	0.42	0.76	
	(0.51)	(0.50)	(0.43)	
HCA training	0.81	1.00	1.00	
	(0.40)	(0.00)	(0.00)	
Observations	37	36	37	37

Notes. Columns 1 - 3 present the mean and standard deviation (in parentheses), for each care home (according to the main treatment). Column 4 lists the p-value for the Mann-Whitney test of equality of distribution for the Bottom-up and Top-dowm care homes, and in parenthesis the p-value of the Kruskal-Wallis test that at least one population distribution of one groups is different from the population distribution of at least another group. *Years of schooling* is the number of years of formal education, excluding professional training. *Employment Years* number of years employed in the job. *Hourly pay* is the hourly pay before taxes, in Euro. *Age* is expressed in years. *Female* is a dummy = 1 if the employee is female. *Native* is a dummy = 1 if the employee was not born or not of Italian ancestry. *HCA training* is a dummy = 1 for employees who completed professional training as healthcare assistants.

Column 4 shows the p-values of the Mann-Whitney test, comparing only the Bottom-up and Top-down facilities, and the Kruskal-Wallis test, comparing simultaneously all three facilities. Both tests reject the hypothesis that the variables *years of schooling* and *hourly pay* are drawn from different distributions. The Mann-Whitney test also rejects the hypothesis that the variable *years of employment* is drawn from different distribution in the Bottom-up and Top-down care homes. The average of *years of employment* (9.69) is, in fact, the same, while in the Control home it is lower (8.3, Kruskal-Wallis significant at 1%). Similarly, *HCA*

training (i.e. professional training as healthcare assistants) is the same for the Bottom-up and Top-Down care homes (all healthcare assistant received some formal training). This analysis shows that workers in the Bottom-up and Top-down care homes have identical average professional profiles in term of years of employment, professional training, and very similar profiles in terms of years of schooling and hourly pay. We can conclude that on they key professional characteristics the two main groups are a very similar and comparable. A group of employees in the Control group does not have any professional HCA training, but inspection of the data shows that they all have at least nine years and half of experience in the job. There are some differences in the demographics and personal characteristics of the workers. The employee of the Bottom-up care home are, on average, moderately younger. In the Top-down care home there is a higher number of workers which were not born in Italy, or of Italian ancestry. Most workers are female in all care homes. These characteristics should be orthogonal to the experimental treatments, but we will control for them in the regressions.

2.2. Workplace details and organization

We study workers in the division of residential services for the elderly within a large non-profit organization. The firm operates in Italy (a country which ranks top across different healthcare systems rankings; WHO, 2000; Bloomberg, 2014), has been established in the market for the last 25 years, employs directly around 1500 workers, and provides services to more than 9000 end-users per year.[14] The 15% workers' turnover of the residential services division is considered very low in the sector.[15] The firm has been be awarded quality control certification on international standards, as well as on management of safety of workers in the workplace. It has also constantly increased its revenues. The residential services division manages 11 facilities for non-self-sufficient and self-sufficient elderly people, for a total of about 750 end-users.[16]

The experiment was located at three care homes, each employing around 40 healthcare workers and hosting 80 residents. The number of residents in the care homes is fixed and

[14]The other core production areas are: home-based services for elderly people; mental health support; disabilities support; child- and minor-care centres.

[15]One of the reasons is that, low turnover levels increase the applicant's score to win public procurements to manage care homes, which is the main source of business for the organization.

[16]Residential facilities are designed for elderly people who do not need 24-hour nursing care, but are unable to live independently. Such facilities may be described as board and care homes, or rest homes. They provide a furnished room, together with all meals, housekeeping and laundry service (see Section 2.3 for further details). They also provide assistance with daily activities such as personal hygiene, dressing, eating, and walking. They are not considered to be medical facilities, but they do have to meet high standards for care and safety.

there is always excess demand for beds, so that no bed is ever left vacant. Each care home is an independently managed organization. Their location is an important feature for the aim of this experiment: the three facilities are remotely located one from each other (approximately, 1-hour drive by car). Personnel from one care home is not acquainted with other care homes' employees and it never happens that there is a swap of workers across facilities.

Each care home is managed by a principal, who reports to the general manager of the residential services area. In each care home there are three personnel units. Each healthcare assistant is assigned to a unit, which is monitored and managed by a unit-specific leader.[17] The unit-leader is an experienced healthcare assistant that was selected by the firm's manager to undertake a specific management path (they are not included as experimental subjects). The three units rely across-the-board on a staff of nurses.

2.3. Experimental subjects: the healthcare assistants

The healthcare assistants will be the experimental subjects. Their job description is defined and limited by law across three main areas:1) *technical:* physical help to the assisted in everyday activities and hygiene (personal hygiene and diuresis, laundry, bedding...); first-aid basic competencies; assistance in administering drugs upon instruction of medical staff (nurses, doctors) or family; check on dietary needs; cooperation with nurses or medical staff to elicit needs; help mobilization of residents (e.g., right position of beds); help in the transport of residents; use of wheelchairs; 2) *social*: healthcare assistants communicate and cooperate with the residents' family to facilitate socialization and relationships; they facilitate socialization among residents; they manage the respect of privacy and ethics. 3) *administrative*: healthcare assistants keep a detailed record of activities and events occurring during working hours; participate in quality-check controls.

The healthcare assistants in the study are all in permanent jobs. The work entails 8-hour shifts, organized on a weekly routine of two mornings, two afternoons, and one night.[18] Nurses and unit-leaders supervise healthcare assistants and continuously check the correct performance of tasks throughout shifts.

[17] The assignments to units is stable, meaning that change of unit in the short-term is uncommon.

[18] The job description of healthcare assistants is standardized by law and a professional one-year qualification may be required to work, which usally entails 1000 hours of courses, including a tutored internship. It is not necessary to hold a high-school degree to be awarded the professional qualification. Healthcare assistants are not nurses and they cannot independently administer drugs.

2.4. Performance and record of mistakes

Healthcare assistants' performance varies mostly in quality. The number of residents to take care of is fixed and the timing of operations is standardized and determined contractually.[19] The key outcome variable of this study will thus be the number of mistakes of healthcare assistants.

We exploit the fact that unit-leaders continuously monitor and carefully record the activities and mistakes of healthcare assistants.[20] The accuracy of such records is guaranteed by two main facts: a) truthful records are fundamental for insurance coverage, and missing or false records are subject to serious disciplinary actions; b) recording facts, operations, medical events is a major component of the healthcare assistant job (as outlined in Section 2.3). The recording of mistakes was based on a pre-set list compiled by the care homes themselves, covering more than 50 items, with additional open fields for special cases. The complete list of the items can be found in Appendix A.8. The detection of mistakes is not discretionary, as there are protocols that objectively determine, e.g., if hygiene was performed correctly or if the resident's aesthetic was taken care of. In Section 5 we will also show that, in the periods before the core experimental treatments, both absolute levels of mistakes and trends are very similar across care homes, which is in turn consistent with similar professional qualification, education and experience across care homes as shown in Table 1.

We consider everyday records during the morning and the afternoon shifts only. This is because operations mainly happen in the morning, and nights are typically uneventful. This is consistent also with experimental records, as the majority of mistakes was recorded in the morning (with few mistakes in the afternoon). In addition, during the night shift, the staffing is minimal and the unit-leader would typically not be present to record mistakes.

3. Behavioural predictions

This section and the following behavioural predictions build on Self-Determination Theory (Deci and Ryan, 2000), in particular on Gagné and Deci (2005), which considers its implications for work motivation. Self-Determination Theory distinguishes between two kinds of motivation: autonomous, when the subject acts with volition and experiences choice, and controlled, when the subject acts as consequence of pressure or coercion. In the context

[19] After discussing the issue with care homes principals and the residential area manager, they have made very clear that the speed of operations is already timed at the frontier of possibilities, that there cannot be a larger number of beds or residents.

[20] In some cases, also nurses and the principal manager recorded mistakes, if they happened to witness one.

of work motivation, SDT hypothesises that a worker's behaviour and well-being depend in a key way on whether incentives are perceived to be autonomous or controlling, possibly leading to sharply different outcomes.

The constructs of controlled motivation and external regulation can be associated with extrinsic incentives (such as rewards, piece-rate pay, or approval whereby actions are driven by the contingency between behaviour and consequences). Standard examples are employees who work and do not shirk only when a supervisor is monitoring their job, or workers who produce more when motivated by monetary, high-powered incentives. Self-Determination Theory, instead, provides a more nuanced insight of extrinsic incentives and proposes that, in some circumstances, they can entail autonomous motivation.[21] *Autonomous extrinsic motivation* is a key component of the following behavioural hypotheses: an employee may find a task unpleasant *per se*, but she can perform it well and with high effort anyway, if that performance is consistent with her own personal goals and values. Gagné and Deci (2005) provide the example of nurses, which is perfectly fitting to the context of this experiment:

> If nurses strongly value their patients' comfort and health and understand the importance of doing their share of the unpleasant tasks for the patients' well-being, the nurses would feel relatively autonomous while performing such tasks (e.g., bathing patients), even though the activities are not intrinsically interesting.

Autonomous extrinsic motivation is different from intrinsic motivation. Intrinsic motivation is driven by the interest in the task itself. With autonomous extrinsic motivation the task is still uninteresting by itself, but it is performed because it is instrumental. The instrumentality, however, is not due just to the contingency (recognition, punishment, reward), but it is associated with personal goals.

In this field experiment, when the management introduces the recognition scheme (a form of external regulation), the experience of choice and participation should trigger autonomous extrinsic motivation. Instead, when the scheme is imposed top-down, workers might frame it as a form of controlled extrinsic motivation.

When considering complex jobs like the ones that we study in this paper, SDT predicts that employees perform better when driven by autonomous motivation, as an improved performance increases the satisfaction in achieving their own goals. This leads to the first prediction.

[21] SDT suggests that there is a continuum from external regulation to internal regulation and analyses different types and degree of internalization. It is not the aim of this paper to distinguish and study the different types of autonomous extrinsic motivation.

Behavioural hypothesis 1 (Bottom-up vs Top-down performance). *The performance in the care home where the new scheme is introduced bottom-up is higher than in the care home where the same scheme is introduced top-down.*

In other words, when employees participate in the choice of some aspects of the scheme, they make the scheme's goals their own, so that they perceive a higher sense of autonomy and therefore increase effort in the task. This does not happen when they do not contribute to the design of the scheme in any way, which decreases autonomy and induces them to frame the scheme as controlling. This behavioural prediction is also consistent with findings in other contexts (for a review see Deci et al. 1999, in the non-profit sector Deckop and Cirka 2000, in higher education Sheldon and Elliot 1998).

This experiment also includes a control group, a care home where the recognition scheme is not introduced throughout the experiment. We consider this care home as a benchmark for performance.

Behavioural hypothesis 2 (Top-down vs Control performance). *After the top-down introduction of the recognition scheme, performance in the top-down care home is inferior to the control group.*

Behavioural hypothesis 3 (Bottom-up vs Control performance). *After the bottom-up introduction of the recognition scheme, performance in the bottom-up care home is no lower than the performance in the control group.*

Prediction (2) entails the idea that workers in the top-down treatment frame the introduction of the recognition scheme as an instrument of control, and thus the extrinsic non-monetary rewards in the scheme do not have an incentivising effect. Rather, they backfire. The mechanism driving Prediction (3) is analogous to Prediction (1).

A final behavioural prediction, which is empirical in nature, concerns the persistence of the effects in predictions (1) - (3). Existing studies on self-determination and performance show that with complex tasks the effect of increased autonomy is persistent (Scheflen et al., 1971), while with mundane tasks it is short lived (e.g., in the context of learning, Grolnick and Ryan 1987). The famous experiment in the context of childcare providers in Gneezy and Rustichini (2000) corroborates the idea that effects may persist after the intervention is removed.

Behavioural hypothesis 4 (Persistence). *The superior performance in the bottom-up care home and the adverse effects on the top-down care home persist even after the recognition scheme has been withdrawn.*

We thus expect some degree of persistence of the effect of participation on motivation, even once the recognition scheme is withdrawn. This would be consistent with the idea that participation allowed the workers to internalise the pursue of enhanced quality as instrumental to their personal goals, and hence they maintain a higher level of quality even subsequently, once the recognition scheme is removed.

4. Experimental design

The experiment ran for three months (14 weeks) and its timeline is divided in four phases (see Figure 1): a 2-week *secret* recording phase; a 3-week overt recording phase, with no incentives attached; a 4-week core treatments phase, with the Bottom-up and Top-down treatments; and a final 5-week overt recording phase, with no incentives attached. Comparing performance in the secret and overt recording phases provides results on Hawthorne-type effects. Comparing the Bottom-up treatment phase with the overt monitoring phase we obtain the main results about the effect of participation on performance (see Section 3, Behavioural hypotheses 1 - 3). Finally, comparing the treatments phase with the last period, where workers return to overt monitoring without incentives, we examine the persistence of the Bottom-up and Top-down treatments' effects (Behavioural hypothesis 4).

We do not randomly assign treatments, rather we exploit the personnel data, logistic information, and other observables to ensure that the three care homes are comparable. To randomise meaningfully one would require exposing workers within the same care home to be in the control group and in the treatment group, which leads to biased estimates of the treatments' effect. Instead, we both exploit the possibility to combine a within-worker analysis with a difference-in-differences analysis. One identification concern could be that the treatment effects may be spurious and driven by unobservable factors occurring at the same time of the treatments. We incorporated in the design and conduct of the whole experiment some measures to address this concern. First, in consultation with the management, the experiment and the treatment phases were run in a period when seasonal changes are typically absent, on the basis of previous years observations. [22] Second, the fact that all three facilities are part of the same organization excludes any organization-related, care-home-specific factors occurring at the same time of the treatments. Third, as we exogenously time and control both the introduction *and* the withdrawal of the scheme, it is highly unlikely that time-varying unobservables drive outcomes precisely both at the time of the introduction *and* of the withdrawal. Finally, besides agreeing with the management that no change other than the treatments happens in the three care homes during the experiment, in Section 7 we provide additional evidence, based on the observed type of observed mistakes, that the treatment effects are consistent with our theory and not driven by other factors.

[22] The management suggested that it would have been better to run the experiment within a season, as typically seasonal and weather changes have impact on the activities and behaviour of end users. For example, in Spring outdoor activities increase. The experiment was thus run in Winter when all activities and operations are indoor.

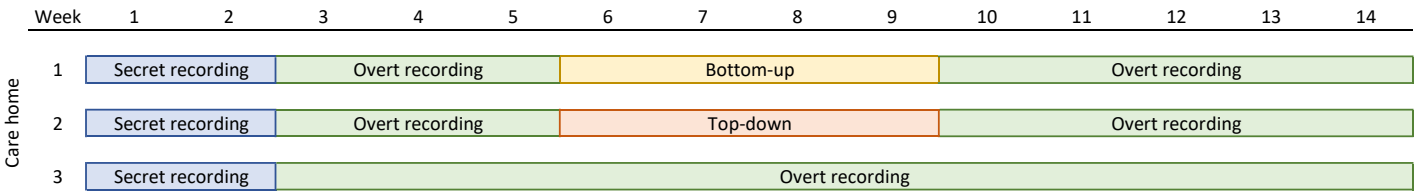


Figure 1: Experimental timeline

We will first present the design of each phase and then the informational structure of the experiment.

4.1. Pre-treatment phases: secret vs overt monitoring and Hawthorne-type effects

We designed the five weeks before the core treatments to record the baseline ability of the employees and control for Hawthorne-type effects. Unit-leaders had always recorded their colleagues' mistakes on block notes they carry in their pockets. This practice has always been common knowledge among employees. However, to create this study's dataset (the number and type of mistakes), we introduced a new paper form to record mistakes systematically and consistently. Potentially, the introduction of a new form may influence the employees performance, even without incentives attached. To control for this Hawthorne-type effect, in the first two weeks the unit-leaders were secretly using the new form to record mistakes, but without informing the healthcare assistants. Importantly, these two weeks also provide us with a baseline record of mistakes, before the introduction of the new form and any other treatments. At the beginning of the third week, the unit-leaders overtly introduced the new form to the other healthcare assistant, making clear that it was not related to any form of incentives/punishment. Comparing the performance in the secret recording period and monitoring period we can test for the presence of Hawthorne-type effects. This also allows us to study the core treatments net of Hawthorne-type effects.

4.2. Treatments phase: the effect of participation on performance

The treatments period ran for four weeks (6 to 9). In the first care home, the area manager invited the employees and the principal manager to a meeting on Wednesday of week 5, where she presented the introduction of the recognition scheme and offered the employees to vote on supposedly secondary features (the "Bottom-up" treatment): its communication format (public or private recognition) and the rule to be used to determine who deserves recognition on the basis of mistakes. In the second care home, the following day (Thursday of week 5) the same area manager performed the same presentation of the introduction of

the same recognition scheme, but without letting the workers vote on any feature (the"Top-down" treatment). The two treatments thus differ only in the possibility for voting on some features of the recognition scheme, but the recognition scheme itself, its presentation format, and the information shared with the employees was exactly the same in the two care homes. All workers in both treatments were informed that the new recognition scheme would have lasted one month as a trial period, after which the management would have evaluated the results and decided whether to introduce it on a permanent basis. The third care home was a control, i.e. no changes were introduced with respect to the overt recording period until the end of the experiment. A detailed description of the two treatments and assemblies follows.

4.2.1. Bottom-up treatment

On Wednesday of week 4 the area manager organized a meeting with all healthcare assistant, the manager, and union representatives in the Bottom-up care home. The area manager seldom meets healthcare assistants and the meeting was thus considered important. During the meeting, the area manager explained that the firm intended to introduce a weekly recognition scheme based on the quality of performance, but that they wanted the healthcare assistants to choose some features of the scheme. The different features were explained in detail, making sure that healthcare assistants understood them, also through a questions and answers session. The menu of alternative features was:

1. Display in the notice board of weekly rankings of all healthcare assistants, sorted by lower mistakes;
2. Display in the notice board of weekly and cumulative rankings of all healthcare assistants, sorted by lower mistakes;
3. Display in the notice board of the names of the top-three healthcare assistants, ranked according to weekly mistakes;
4. Display in the notice board of the names of the top-three healthcare assistants, ranked according to weekly and cumulative rankings;
5. Private one-to-one meeting with the principal manager for the top healthcare assistant(s) based on weekly rankings;
6. Private one-to-one meeting with principal manager for the top healthcare assistant(s) based on weekly and cumulative rankings.

By "weekly" mistakes we mean the ranking according to the sum of a week's mistakes of each employee, while cumulative rankings sum also all previous weeks' mistakes. The collectively chosen incentive was n. 5.[23]

[23]This recognition scheme and the metric used was considered fair by the participants and also by union

4.2.2. Top-down treatment

On Thursday of week 5, the area manager organized a meeting with all healthcare assistants in the Top-down care home. The meeting was analogous to the Bottom-up treatment, with the only difference that the area manager communicated to the healthcare assistants that from week 5 a recognition scheme (n.5 above) was to be introduced. The area manager explained the same details of the scheme as above, but did not mention the other possibilities nor that it was a scheme chosen by another care home's healthcare assistants, nor that it was implemented elsewhere.

4.3. Persistence phase

The post-treatment period lasted from week 10 to week 14. During this period, for all care homes the recognition scheme was withdrawn, returning to the only overt monitoring of mistakes (thus returning to the same condition as in weeks 3 - 5). The purpose of this treatment is to check the persistence of any possible effect of the recognition scheme and to control whether the healthcare assistants behaviour respond both to its introduction and its withdrawal.

4.4. Informational structure

The nature of the experiment and its detailed ends were known only to the residential area manager. Even the principal managers were only broadly aware of the collaboration with the university, but did not know the exact aim of the project, and they never met each other before and during the experiment, so they were not aware of what was happening in the other structures. The residential area manager also made sure that no information leaked to unit-leaders and healthcare assistants about the project and its aims. Unit-leaders were progressively informed with some notice of the introduction of the mistakes secret and overt record keeping, and of the introduction of the recognition scheme, but were instructed by the principal not to leak the information to other healthcare assistants, until the area manager would summon the explanatory meeting with them. They were as well not informed of what was happening in other houses.[24] The perception of unit-leaders and healthcare assistant was that the firm's management wanted to introduce a new protocol and have a trial period

representatives that attended the meeting. If there were ties in best performance, the principal would have to congratulate (privately and separately in her office) all employees in the tie.

[24]A point, which maybe worth discussing, is that during the treatments an increase in the recording might not be to due to an increase in mistakes, rather to a change in the attitude of unit-leaders in terms of strictness and precision of recording. This was not the case, as explained in Section 2.4, because of the consequences of misreporting information, which in some cases may be of legal nature. In any case, since unit-leaders know in advance the introduction of new protocols, if there is an "empowerment" effects on unit-leaders we should observe it since the first weeks of the experiment and in comparison with the control

of it. The introduction and test of new protocols is common practice and therefore was not perceived as unusual within the care homes.

5. Descriptive statistics

Figure 2: Total weekly mistakes by care home and experimental period.

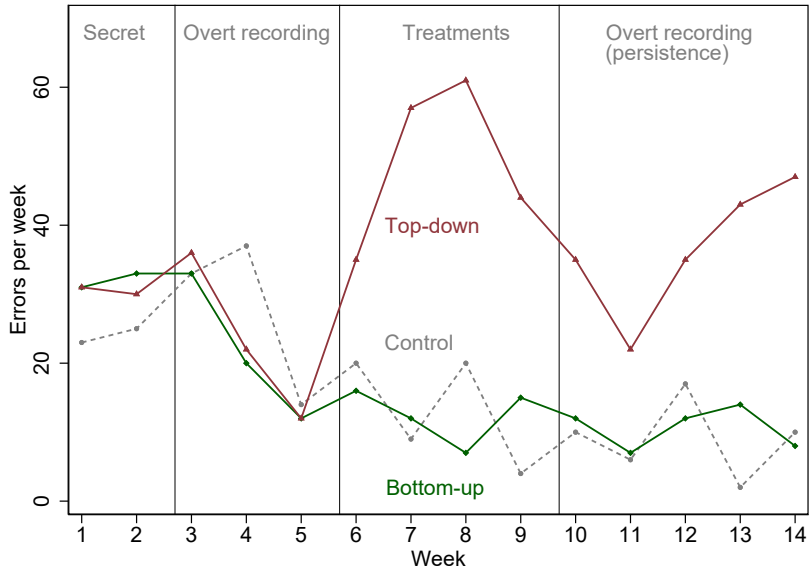


Figure 2 shows total weekly errors for each care home across the experimental weeks and phases. We refer to it to provide a qualitative overview of the main results of the experiment, which we present using econometric methods in the Sections 6.1 and 7. The three care homes record a similar magnitude of mistakes in the secret recording phase. The number of mistakes in the Control group is slightly lower but, importantly, it is very similar for the Bottom-up and Top-down care homes. Table 2 also confirms the similarity in recorded mistakes in the first five weeks before the introduction of the core treatments, showing that the average weekly mistakes per worker is very similar across the care homes. Table 3 further shows that the two-sample Kolmogorov-Smirnov test does not reject the hypothesis of equality of distribution for all pairs of care homes in the secret and overt recording phases. The equality of the distributions of mistakes in the first five weeks provides evidence that recording standards are very similar and comparable, if not identical, across the care homes.

group. We anticipate that there is no evidence of this effect on unit-leaders on the basis of the analysis of results. Information gathered through follow-up meetings with the care-home principals confirm also that this "empowerment" effect was not observed.

Figure 2 shows that the introduction of overt recording in week 3, without any incentive attached, eventually reduces the number of mistakes in all care homes (with a one-week lag in the Control care home). In the Control group, mistakes remain at a similar level after week 4, with a slight downward trend, until the end of the experiment. This evidence supports the presence of Hawthorne-type effects, which we will discuss in Section 6.2.1. Considering the pre-treatment periods, we can also conclude that the parallel trends assumption, key for the difference-in-difference analysis to follow, is verified. In addition, also absolute overall mistakes and mistakes per worker are at the same level across care homes in the first five weeks.

In the treatments phase, the graph shows this paper's main result: mistakes in the Top-down care home increase sharply well above the level of the secret and overt monitoring phases, unlike in the Bottom-up and Control homes. It is difficult to detect a clear effect of the Bottom-up treatment in comparison with the control group. This is possibly due to the fact that in both homes the mistakes almost reach a zero-lower-bound (see Table 2, around 0.35 weekly mistakes/worker, on average for each care home). Finally, Figure 2 shows clear evidence of the persistence of the effects after the recognition system was suspended: in the Top-down care home, the level of mistakes reduces, but remains at higher level than the other care homes, and also higher than in the pre-treatment phases. There is some further decrease in mistakes in the Bottom-up and Control care homes, but suspending the recognition scheme does not seem to increase mistakes back to the pre-treatment's levels.

Figure 3 presents the kernel density estimates of the distribution of weekly mistakes across experimental phases and care homes. The red line is associated with the Top-down care home, the green line with the Bottom-up home, the blue line with the Control care home. Each care home's mistakes are not normally distributed, with the largest share of observations recording a few or no mistakes. The distributions almost coincide in the pre-treatment phases, and for the Bottom-up and Control care homes continue to overlap throughout. Considering panel (c) and (d), one can see that the dispersion increases in the Top-down care home and that the distribution is similar in both phases, possibly with the lack of a right tail in the persistence period. The findings in Figure 3 are confirmed in Table 3, which shows Kolmogorov-Smirnov tests of equality of distribution between care homes pairs. The test rejects with p-value at least at 1% that the distribution of the Top-down mistakes is equal to either of the other care homes, both in the treatments and in the persistence phases. The distribution of mistakes in the Bottom-up and Control homes seems to be equal across all phases.

In summary, this Section has shown that the recording of mistakes has been conducted at the same standard across care homes. It has also presented evidence that the parallel

Table 2: Weekly mistakes per worker

	(1) Control	(2) Bottom-up	(3) Top-down
Pre-treatment (5 weeks)	0.68	0.70	0.71
Treatment (4 weeks)	0.34	0.34	1.33
Persistence (5 weeks)	0.23	0.29	0.98
Total working hours	14616	14278	14392
Total number of workers	39	37	37

Notes. The first three row represent the average number of weekly mistakes per worker in each care home, by period. *Total working hours* is the total number of hours worked by the healthcare assistants for which unit-leaders monitored mistakes. *Total number of workers* is the number of experimental subjects, which excludes unit-leaders and occasional healthcare assistants.

trend assumption is likely to be fulfilled, as indeed even in the Bottom-up care home the trend continues to be the same as in the control treatment for the entire 14 weeks of the experiment. Together with the balance of observables shown in Section 2.1, the fulfilment of this assumption is important for the validity of the difference-in-difference analysis which we present in the following section.

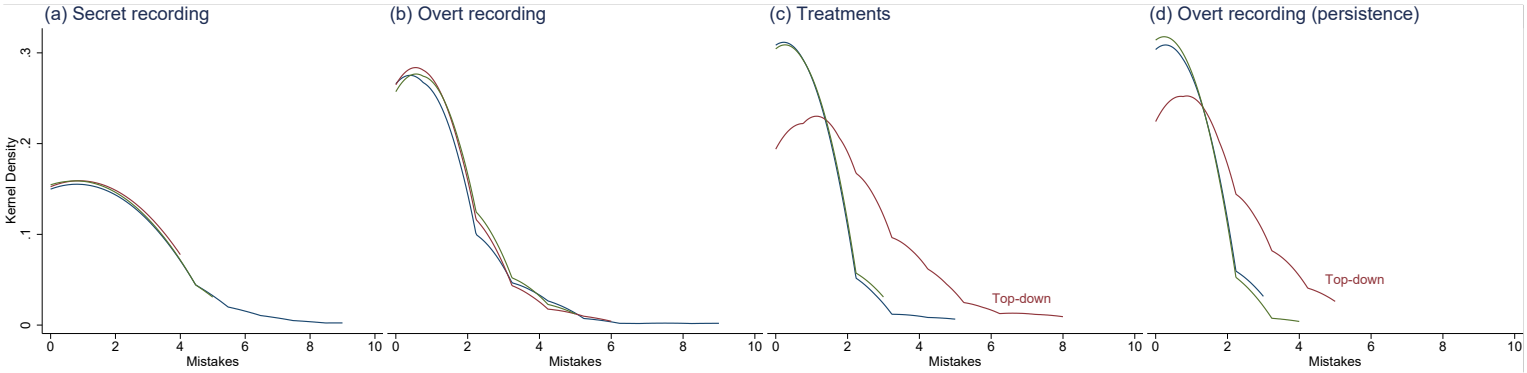
Table 3: Kolmogorov-Smirnov test, exact p-values.

	Secret recording	Overt recording	Treatments	Persistence
Top-down/Bottom-up	0.959	0.976	0.000	0.000
Bottom-up/Control	1.000	0.238	1.000	0.981
Top-down/Control	0.965	0.742	0.000	0.000

6. Empirical strategy and results

In this Section, we use econometric methods to test the behavioural predictions in Section 3. In the first subsection we will test the main hypothesis of this paper: a bottom-up process

Figure 3: Distribution of mistakes by phase



reduces mistakes in comparison with top-down rules. We will focus on the Bottom-up and Top-down care homes only, which employ the same recognition scheme, but differ in the way in which the recognition scheme was introduced (Behavioural hypothesis 1). In the second subsection, we examine the effect of the Bottom-up and Top-down treatments in comparison with the Control care home (Behavioural hypotheses 2 and 3). In the last subsection we provide some evidence of Hawthorne-type effects following the introduction of the overt recording of mistakes.

6.1. The effect of participation vs. the Top-down treatment

As it is apparent from Figure 2 and the descriptive statistics in Section 5, in the Top-down care home a higher number of mistakes was recorded than in the Bottom-up care home. Using the Top-down care home as baseline, we study the effect of the participation treatment (Behavioural hypothesis 1) and its persistence (Behavioural hypothesis 4). We estimate the following difference-in-difference model:

$$m_{i,u,t} = \kappa + \alpha_0 T + \alpha_1 PT + \beta U + \gamma (U \cdot T) + \delta (U \cdot PT) + \lambda_u + \mu H_{i,t} + \theta X_i + \varepsilon_{i,u,t} \quad (1)$$

where $m_{i,u,t}$ is the weekly number of mistakes of healthcare assistant i working in unit $u = 1, ..., 6$ in week $t = 3, .., 14$; T and PT are time dummies equal to one in the treatments phase and the persistence phase, respectively; U is an indicator for observations in the Bottom-Up care home; and $\cdot$ denotes an interaction between dummy variables; λ are unit-level fixed effects to control for unobserved permanent differences in mistakes across units; error terms $\varepsilon_{i,u,t}$ are clustered at individual level (as individual mistakes are likely to be correlated across weeks);[25] $H_{i,t}$ is the number of hours actually worked by the healthcare

[25] We consider also a model with unit-week fixed effects, to control for unobserved differences across units

assistant i in week t; X_i is a set of individual, time-invariant control variables including years of employment, years of schooling, age in years, sex, native status, and hourly pay.

The coefficients of interest are γ, the difference-in-difference estimator which measures the expected average difference in mistakes due to the Bottom-up treatment compared to the Top-down treatment; and δ, the average difference in mistakes due to the Bottom-up treatment in the post-treatment phase when the recognition scheme is withdrawn in both care homes, i.e. the persistence of the Bottom-up treatment effect.

Table 4 presents the main results of the paper. All specifications, with significance at 0.001%, show that the Bottom-up treatment reduces one mistake per week per worker compared to the Top-down treatment. At the aggregate care-home level, the Bottom-up treatment thus delivers around 40 mistakes per week less than the Top-down treatment, with considerable implications for the management of the healthcare assistants' operations. This result supports Behavioural hypothesis 1.

The second coefficient (*Persistence*, -0.759, significant at 0.001%) is lower. This suggests that healthcare assistants respond not only to the introduction, but also to the withdrawal of the scheme. Together with previous results, this provides confidence that workers react to the experimental manipulation, rather than to other unobservable factors. The finding suggests that, thought smaller, the effect of the treatment persists even after the recognition scheme is withdrawn.[26] These results are consistent with the descriptive statistics in Table 2. Overall, these results confirm Behavioural hypothesis 4.

In the pre-treatment period, which includes also the secret recording weeks, average mistakes per worker are equal to around 0.7 in both care homes. Mistakes almost double in the Top-down care home and halve in the Bottom-up care home in the treatments phase. After the withdrawal of the recognition scheme, mistakes remain higher than in the pre-treatment period only in the Top-down care home.

Including controls in Column (2) leads to a negligible reduction in the coefficients of interest and in their heteroschedasticity-robust standard errors. *Hourly pay* has a small positive effect, which is not significant in some of the models (and never below the 5% level). The coefficients of other control variables are all close to zero. Except *Working hours* and *Schooling years* in some models, these coefficients are not significant.

which are specific to some weeks. We show also results with standard errors clustered at unit-week level, to account for the possibility that observations in a specific week are correlated within a unit. Clustering by unit only may be inappropriate because of the small number of clusters, but leads to results with similar significance.

[26] If compared with the secret recording weeks only (i.e. before any overt change is introduced), the coefficient is -0.84 and significant. This measures the overall post-treatment effect, including the continuation of the overt recording of mistakes.

Table 4: The effect of participation: Bottom-up vs Top-down

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Weekly mistakes	OLS	OLS	OLS	OLS	OLS	GLS
Bottom-up	-1.054^{****}	-1.053^{****}	-1.071^{****}	-1.071^{****}	-1.071^{****}	-1.083^{****}
	(0.220)	(0.213)	(0.209)	(0.206)	(0.259)	(0.198)
Persistence	-0.759^{****}	-0.740^{****}	-0.757^{****}	-0.757^{****}	-0.757^{***}	-0.767^{****}
	(0.190)	(0.187)	(0.181)	(0.185)	(0.262)	(0.177)
Working hours		0.019^{****}	0.020^{****}	0.020^{****}	0.020^{****}	0.024^{****}
		(0.004)	(0.003)	(0.004)	(0.003)	(0.004)
Employment (years)		-0.005	-0.032^{**}	-0.032^{*}	-0.032^{*}	-0.006
		(0.015)	(0.016)	(0.018)	(0.016)	(0.023)
Schooling (years)		-0.042^{**}	-0.050^{***}	-0.050^{**}	-0.050^{***}	-0.045
		(0.017)	(0.016)	(0.022)	(0.015)	(0.031)
Native		0.011	-0.027	-0.027	-0.027	-0.034
		(0.086)	(0.085)	(0.104)	(0.095)	(0.138)
Female		0.019	-0.043	-0.043	-0.043	0.034
		(0.117)	(0.118)	(0.146)	(0.091)	(0.151)
Age (years)		-0.006	-0.002	-0.002	-0.002	-0.006
		(0.004)	(0.005)	(0.006)	(0.004)	(0.007)
Hourly pay		0.231	0.368^{**}	0.368^{*}	0.368^{**}	0.201
		(0.161)	(0.159)	(0.193)	(0.142)	(0.264)
Adjusted R^2	0.113	0.143	0.207	0.207	0.207	
Overall R^2						0.155
Fixed effects	NO	NO	YES	YES	YES	NO
Clustered SE	NO	NO	NO	YES	YES	YES
Observations	807	807	807	807	807	807

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes. Dependent Variable: Weekly individual mistakes; (1) OLS with robust standard errors, without controls; (2) OLS with robust standard errors; (3) OLS with robust standard errors and unit-level fixed effects; (4) OLS with cluster-robust standard errors at individual level, and unit-level fixed effects; (5) OLS with cluster-robust standard errors at unit-week level and unit-level fixed effects; (6) GLS with random effects and clustered-robust standard errors at individual level. Clustered standard errors are computed over 9*12 clusters at unit-week level, and 73 clusters at individual level. *Bottom-up* is an interaction dummy between time dummy for the treatment phase (not shown in the table) and the Bottom-up care home dummy. *Persistence* is an interaction dummy between time dummy for the post-treatment phase (not shown in the table) and the Bottom-up care home dummy. *Working hours*, effective weekly working hours. Time-invariant controls: *Schooling (years)*, number of years of formal education, excluding professional training. *Employment (years)*, number of years employed in the job. *Hourly pay* is the hourly pay before taxes, in Euro. *Age* is expressed in years. *Female* is a dummy = 1 if the employee is female. *Native* is a dummy = 1 if the employee was not born or not of Italian ancestry. Data include weeks 3-14.

Columns (3) - (5) test the robustness of the model significance including unit-level fixed effects, which control for permanent differences in the units, while changing the clustering of standard errors. The introduction of fixed effects in columns (3) - (5) leads to a very small increase in the coefficients of interest. To account for correlations of mistakes for the same workers, in Column (4) standard errors are clustered at individual level. This leads to a small increase in standard errors, with the exception of the *Persistence* coefficient. In Column (5), clustering at unit-week level increases standard errors, but does not affect the overall significance of the main coefficients of interest (with δ still significant at 0.01%). Finally, in Column (6) we estimate a random effects model with cluster-robust standard errors at individual level, which yields similar coefficients estimates and lower standard errors than the other models.[27]

Summing up, in this section we have shown the main result of the paper, which suggests that introducing top-down the recognition scheme generated a sizeable increase in the number of mistakes. This result, which supports Behavioural hypothesis 1, highlights the importance of adopting bottom-up procedures to avoid the scheme to backfire. This is a possible explanation of mixed results in the labour economics literature on recognition schemes (e.g. Ashraf et al. 2014, Bradler et al. 2016), which might be reconciled when accounting for different degrees of participation and in the way the scheme was introduced in the context of those and future studies.

6.2. Participation and Top-down vs. Control

In this section we test Behavioural hypotheses 3 and 2, comparing to the Control care home the effect of the Bottom-up treatment and, distinctly, of the Top-down treatment. Using the following econometric difference-in-difference models, we first test the effect and persistence of the Bottom-up treatment compared to the Control care home:

$$m_{i,u,t} = \kappa + \alpha_0 T + \alpha_1 PT + \beta\, U + \gamma\,(U \cdot T) + \delta\,(U \cdot PT) + \lambda_u + \mu H_{i,t} + \theta X_i + \varepsilon_{i,u,t} \quad (2)$$

and then the effect and persistence of the Top-down treatment compared to the Control care home:

[27] As the dependent variable is not normally distributed, one concern may be that ordinary least squares regressions may be biased. In the Appendix Table (A.9) we show that the results are similar using Poisson estimates. Using zero-inflated negative binomial models to account for excess zeros in the dependent variable yield similar results. Adding a linear individual trend and unit-week fixed effects do no affect the sign and significance level of the coefficient. Qualitatively similar and robust results can also be obtained using weekly mistake/hour as dependent variable. These results are available upon request.

Table 5: The effect and persistence of treatments vs the Control care home

	(1)	(2)	(3)	(4)
Dep. var.: Weekly mistakes	OLS	OLS	OLS	OLS
Bottom-up	0.155	0.157		
	(0.180)	(0.176)		
Pers. Bottom-U.	0.218	0.241		
	(0.167)	(0.162)		
Top-down			1.221****	1.234****
			(0.219)	(0.207)
Persistence Top-down			0.990****	1.034****
			(0.182)	(0.175)
R^2	0.078	0.113	0.197	0.234
Controls	NO	YES	NO	YES
Fixed effects	YES	YES	YES	YES
Clustered SE		YES	YES	YES
Observations	808	808	815	815

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes. Dependent Variable: Weekly individual mistakes; Clustered standard errors at individual level (73 and 74 clusters). Unit-level fixed effects. *Bottom-up* is an interaction dummy between time dummy for the treatment phase (not shown in the table) and the Bottom-up treatment dummy. *Pers. Bottom-U.* is an interaction dummy between the time dummy for the post-treatment phase (not shown in the table) and the Bottom-up care home dummy. *Top-down* is an interaction dummy between time dummy for the treatment phase (not shown in the table) and the Top-down care home dummy. Controls: *Hours* effective weekly working hours. *Persistence Top-down* is an interaction dummy between the time dummy for the post-treatment phase (not shown in the table) and the Top-down care home dummy. Time-invariant controls: *Schooling (years)*, number of years of formal education, excluding professional training. *Employment (years)*, number of years employed in the job. *Hourly pay* is the hourly pay before taxes, in Euro. *Age* is expressed in years. *Female* is a dummy = 1 if the employee is female. *Native* is a dummy = 1 if the employee was not born or not of Italian ancestry. Data include weeks 3-14.

$$m_{i,u,t} = \kappa + \alpha_0 T + \alpha_1 PT + \beta D + \gamma\left(D \cdot T\right) + \delta\left(D \cdot PT\right) + \lambda_u + \mu H_{i,t} + \theta X_i + \varepsilon_{i,u,t} \quad (3)$$

where $m_{i,u,t}$ is the number of mistakes of healthcare assistant i in week $t = 3, .., 14$, working in unit $u = 1, ..., 6$ in model (2) and $u = 4, ..., 9$ in model (4); T and PT are time dummies equal to one in the treatments phase and the persistence phase, respectively; U is equal to one for observations in the Bottom-Up care home and D is equal to one in the top-Down care home; $\cdot$ denotes an interaction between dummy variables; λ are unit-level fixed effects to control for unobserved permanent differences in mistakes across units; errors $\varepsilon_{i,u,t}$ are clustered at individual level (as individual mistakes are likely to be correlated across

weeks); $H_{i,t}$ is the number of hours actually worked by the healthcare assistant i in week t; X_i is a set of individual, time-invariant control variables including years of employment, years of schooling, age in years, sex, native status, and hourly pay.

Table 5 presents the results. Columns (1) and (2) show that the effect on mistakes of the Bottom-up treatment compared to the control group is small, positive and not significant at conventional levels.[28] This implies that the the recognition scheme, in conjunction with the bottom-up process, did not have any significant effect on mistakes. This result does not seem to support Behavioural Hypothesis 3. An alternative explanation is that the sole introduction of overt recording reduced mistakes to their natural lower-bound, i.e. near zero, and hence the effect cannot be detected.

Columns (3) and (4) show that, compared to the Control group, the Top-down treatment induces an increase of around 1.2 mistakes per week, and that this effect, though slightly smaller, is persistent after suspending the recognition scheme. Together with the descriptive statistics in Section 6.1, this result is key. The findings confirm, in this context, that recognition without participation backfires compared to bottom-up processes *and* to the Control group (Behavioural hypothesis 2).

6.2.1. Hawthorne-type effects

In this section we will briefly discuss the presence of Hawthorne-type effects, which may have been generated by the introduction of a new sheet to record mistakes, even though without any incentive attached in weeks 3 to 5. We staggered the overt recording phase and the treatments phase to disentangle the effect of introducing a new recording sheet from the effect of the recognition scheme, top-down and bottom-up processes themselves. As Hawthorne-type effects were not the focus of the design of the experiment, we can only perform a before-after analysis. In Figure 2 it appears that, after introducing the new sheet in Week 3, mistakes reduce in all homes by the end of Week 5. However, the reduction is not immediate: it can be observed with a one-week lag for the Bottom-up and Top-down care homes, and with a two-week lag in the Control care home. We test the presence of Hawthorne effects using the following econometric models, which includes a one-week lag for the effect to take place:

$$m_{i,u,t} = \kappa + \alpha\, HT_{+1} + \lambda_u + \mu H_{i,t} + \theta X_i + \varepsilon_{i,u,t} \tag{4}$$

for all care homes in weeks 1 to 5, and for the sole Control care home in weeks 1 to 14; $m_{i,u,t}$ is the number of mistakes of healthcare assistant i in week t, working in unit u; HT_{+1}

[28] Controls have similar coefficients (close to zero) and significance as in Table 4.

Table 6: Hawthorne-type effects

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Weekly mistakes						
(lag) overt	-0.312^{***}		-0.459^{****}			
monitoring	(0.095)		(0.103)			
Overt monitoring		-0.128		-0.307^{**}	-0.310^{**}	-0.310^{**}
		(0.097)		(0.129)	(0.131)	(0.128)
Adjusted R^2	0.109	0.094	0.121	0.039	0.089	0.089
Controls	YES	YES	YES	YES	YES	YES
Fixed effects	YES	YES	YES	NO	YES	YES
Clustered SE	YES	YES	YES	YES	YES	YES
Observations	504	504	470	470	470	470

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes. Dependent Variable: Weekly individual mistakes. Ordinary Least squares regressions. Column (1) - (2) include data for all care homes in weeks 1 to 6. Columns (3) to (6) include data for all weeks for the Control group only. Clustered standard errors at individual level in Columns 1 and 2 (110 clusters), 3 - 5 (37 clusters), at unit-week level in Column 6 (42 clusters). Unit-level fixed effects. *(Lag) Overt monitoring* is an indicator for weeks 4 - 14; *Over monitoring* is a time dummy =1 for all weeks after overt recording is first introduced (weeks 3 - 14). Controls include: *Working hours*, effective weekly working hours. Time-invariant controls: *Schooling (years)*, number of years of formal education, excluding professional training. *Employment (years)*, number of years employed in the job. *Hourly pay* is the hourly pay before taxes, in Euro. *Age* is expressed in years. *Female* is a dummy = 1 if the employee is female. *Native* is a dummy = 1 if the employee was not born or not of Italian ancestry.

is an indicator for weeks 4 to 14; we also test the same model using the effective week of the introduction of overt monitoring HT; the rest of the notation is the same as in the previous sections.[29]

In Table 6, including all care homes in weeks 1 to 5 and unit-level fixed effects, Column (1) shows that there is a negative effect (lagged of one week) which is significant at 0.01%. Using the non-lagged indicator for the Overt-recording phase, Column (2) shows that there is a negative effect on mistakes, which however is not significant with standard errors clustered at individual level. This is somehow unsurprising, given the consideration about the descriptive statistics at the beginning of this section. In Columns (3) - (6) we consider only the Control care home in weeks 1 to 14. In Column (3), the coefficient implies a reduction of around

[29] λ are unit-level fixed effects to control for unobserved permanent differences in mistakes across units; $\varepsilon_{i,u,t}$ are clustered at individual level; $H_{i,t}$ is the number of hours actually worked by the healthcare assistant i in week t; X_i is a set of individual, time-invariant control variables including years of employment, years of schooling, age in years, sex, native status, and hourly pay.

0.5 mistakes per worker and significant at 0.01%, with unit fixed effects and standard errors clustered at individual level. Using the non-lagged indicator, the coefficient is negative and significant at 5% level across combinations of fixed effects and standard errors' clustering, implying a reduction of 0.3 mistakes per worker. These compare to the Control care home's average of around 0.8 in the secret phase (i.e., a reduction of more than 50% when we consider a lagged effect), which is consistent with Figure 2.

One can speculate that, rather than Hawthorne-type effects, the decrease in mistakes is due to an anticipatory effect, despite the workers being aware that there was no remuneration or recognition attached. Were that case, in the Control care home after a number of weeks since the introduction of the overt recording mistakes should have increased again, once it would have been clear to the workers that their expectations were incorrect. Evidence from Figure 2 shows that this is not the case: in the Control care home the effect of overt recording does not fade until the final weeks of the experiment. On the contrary, there is a slight downward trend. This evidence points to the presence of genuine Hawthorne-type effects, rather than anticipatory effects.

7. Why a top-down recognition scheme backfires

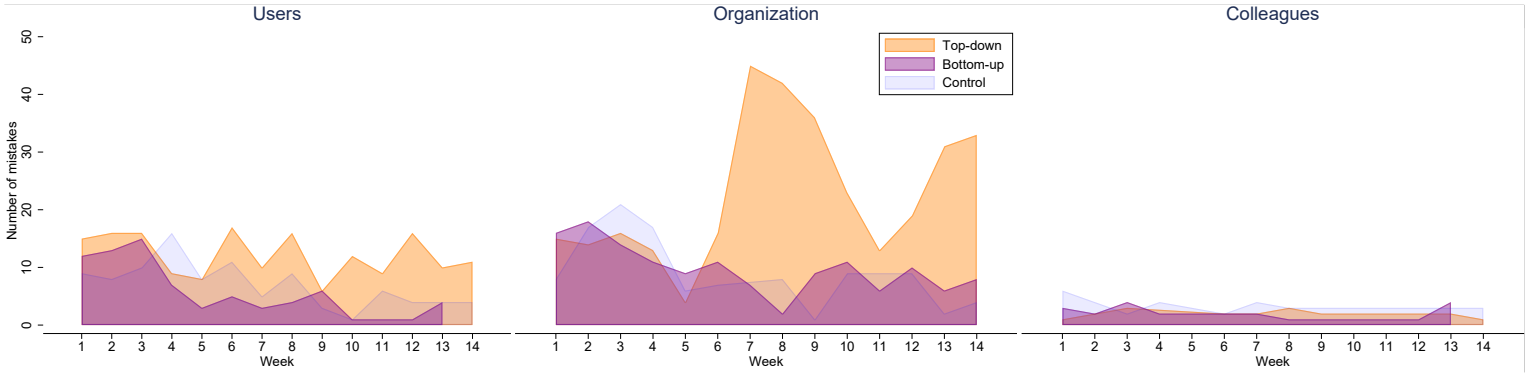
To provide a more nuanced insight of the workers behaviour and clues about their possible drivers, information about the precise type and nature of each mistake was collected during the experiment. In this section, using these data we present and discuss evidence on whether the treatments affect specific types of mistakes or cause an increase across the board. We sorted mistakes into three broad categories, according to the subjects who are directly affected by the mistakes: the end-users (the elderly), the organization (protocols, processes, and management), or colleagues.[30] Figure 4 shows the mistakes count by type along the 14 weeks of the experiment. At glance, it can be seen that the overall increase in mistakes in the Top-down care home is driven by a sharp increase in the mistakes belonging to the *Organization* type, well beyond the level of the secret recording phase.[31] *Users* mistakes are also higher in the Top-down care home, too, though still lower than secret phase's levels.

To test the findings suggested by these descriptive statistics, we first consider the Bottom-up and Top-down care homes only, and estimate the econometric model in equation (1) in Section 6.1 distinctly for each of the three broad types of mistakes, the dependent variable being the weekly number of mistakes of that type. Table 7 presents the results of these estimates in Columns (1) - (3). Considering the *Bottom-up* coefficient, the mistakes in the

[30] See Appendix A.8 for a detailed breakdown.

[31] There was no other event, other than the introduction of the recognition scheme, which may have caused the increase in organization mistakes.

Figure 4: Types of mistakes



Organization column largely account for the overall reduction in mistakes, compared to the Top-down care home. For the other types, the coefficient is smaller and not significant. After the recognition scheme was discontinued, the *Persistence Bottom-up* coefficient shows that the Top-down care home witnesses a reduction in *Organization* mistakes (significant at 0.01%) and an increase in *Users* mistakes (significant at 5%).

To double-check the robustness of these results, we test whether they are consistent when we compare the Top-down care home and the Control group. We also test whether similar findings can be obtained comparing the Bottom-up care home and the Control, or if they are specific to the Top-down treatment. We estimate the following model distinctly for each broad type of mistakes:

$$m_{i,u,t,e} = c + \alpha_0 T + \alpha_1 PT + \beta_0 U + \gamma_0 (U \cdot T) + \delta_0 (U \cdot PT) + \\ + \beta_1 D + \gamma_1 (D \cdot T) + \delta_1 (D \cdot PT) + \lambda_u + \mu H_{i,t} + \theta X_i + \varepsilon_{i,u,t} \tag{5}$$

where $m_{i,u,t,e}$ is the weekly number of type-e mistakes of healthcare assistant i working in unit $u = 1, ..., 6$ in week $t = 3, .., 14$; T and PT are time dummies equal to one in the treatments phase and the persistence phase, respectively; U is equal to one for observations in the Bottom-Up care home and D is equal to one in the top-Down care home; λ are unit-level fixed effects to control for unobserved permanent differences in mistakes across units; error terms $\varepsilon_{i,u,t}$ are clustered at individual level.

Columns (2) - (4) in Table 7 report the results. Compared to the Control group, the effect of the Top-down treatment is qualitatively similar to columns (1) -(3), i.e. there is a significant increase in *Organization* mistakes only. In the persistence phase, *Organization*

Table 7: The effect of participation on different types of mistakes

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: Weekly mistakes	Users	Organization	Colleagues	Users	Organization	Colleagues
Bottom-up	−0.101	−0.639****	−0.061	0.004	0.165*	−0.041
	(0.082)	(0.099)	(0.040)	(0.078)	(0.093)	(0.034)
Pers. Bottom-up	−0.193**	−0.375****	−0.009	0.024	0.190*	−0.006
	(0.093)	(0.103)	(0.030)	(0.078)	(0.098)	(0.031)
Top-down				0.106	0.804****	0.019
				(0.090)	(0.109)	(0.040)
Persistence Top-down				0.218**	0.565****	0.003
				(0.095)	(0.101)	(0.027)
Adjusted R^2	0.072	0.093	0.001	0.062	0.107	0.012
Fixed effects	YES	YES	YES	YES	YES	YES
Clustered SE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES
Observations	898	898	898	1335	1335	1335

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes. Dependent Variable: Weekly individual mistakes, by type; Columns (1) - (3) include data for the Bottom-up and Top-down care homes only. Columns (4) - (6) for all care homes. Ordinary Least Squares regressions with cluster-robust standard errors at individual level (73 and 74 clusters) and unit-level fixed-effects. *Bottom-up* is an interaction dummy between time dummy for the treatment period (not shown in the table) and the Bottom-up treatment dummy. *Pers. Bottom-up* is an interaction dummy between time dummy for the post-treatment period (not shown in the table) and the Bottom-up treatment dummy. Controls include: *Hours*, *Schooling (years)*, *Employment (years)*, *Hourly pay*, *Age*, *Female*, *Native*. Data include weeks 3-14.

mistakes decrease and *Users* mistakes increase and are significant at 5%. Altogether, the Top-down treatment effect is consistent considering different baseline groups (the Bottom-up care home and the Control care home), and thus would suggest that the increase in mistakes targets the organization, rather than users. When the recognition scheme is suspended, the Top-down employees 're-balance' qualitatively the type of mistakes, reducing mistakes towards the organization and increasing mistakes towards the users.

This effect cannot be observed in the Bottom-up care home. Comparing the Bottom-up care home with the Control care home, the treatment and persistence coefficients are small and positive only for the *Organization* mistakes, but their magnitude is similar and their significance is mild (only at 10% level). For the other types of mistakes, the coefficients are close to zero and not significant, with high standard errors, and there is no significant change in the persistence phase. This 'placebo' comparison brings some evidence that the

effect on mistakes' type in the Top-down care home is due to the treatment itself and cannot be observed in the Control or Bottom-up care homes.[32]

Overall, we can state that in the Top-down treatment workers tend to be relatively less motivated to exert effort in the domain of mistakes related to the organizational rules. Ultimately, managers are responsible for the introduction of the recognition scheme, and the highest level of disengagement is associated precisely in instances where tasks pertain to organizational and managerial-related tasks. This is likely because the scheme is perceived as controlling (more monitoring, more demanding performance). While the scheme is in place, employees may perceive rules associated with those tasks as controlling and external, unlike rules associated with tasks which strictly impact the elderly. Once the scheme is suspended, the mistakes' type realigns with the other care homes.

As mentioned in Section 3, there might be additional or alternative explanations. We considered whether reciprocity is a credible candidate driver: as managers did not consult workers, workers' do not facilitate the managers' tasks, in return. We would also expect positive reciprocation in the Bottom-up treatment, however. Instead, the evidence in the Bottom-up care home does not seem to be consistent with a reciprocity-based explanation. In the Bottom-up care home we did not observe an improvement in areas that affect the organization and the management. In fact, if anything, Table 7 shows that there is a small increase in *Organization* mistakes also in the Bottom-up care home, compared to the Control. These findings suggest that the autonomy-control mechanism is the most likely or dominant driver of the backfiring of the scheme in the Top-down care home.

8. Conclusions

In this paper we designed a natural field experiment to study the effect of workers' participation on the quality of performance. To this aim, we introduced a recognition scheme following two distinct procedures: a top-down implementation and a bottom-up procedure. We showed that workers in the Top-down treatment make more mistakes with respect to those in the Bottom-up treatment, and that this effect persists also after the withdrawal of the recognition scheme. This result suggests that non-monetary recognition schemes may backfire if introduced without participation, and confirms the positive impact of participation on effort and performance. When the quality of performance decreases, this is due mainly to mistakes related to the organization's processes and protocols. By contrast,

[32] Together with the facts that mistakes respond to both the introduction and the withdrawal of the scheme, and that no other scheme or change is running at the same time, it also provides further evidence that the behavioural change is driven by the treatment and not by other unobservable factors.

this type of 'selective" change in performance cannot be observed in the Bottom-up care home, suggesting that, rather than a form of reciprocity-retaliation, the employees exhibit more disengagement with management-specific tasks, more in line with self-determination theory.

The Bottom-up treatment does not display a significant reduction in mistakes with respect to the Control group. Nevertheless, the negative outcome in the Top-down treatment suggests that implementing participatory processes is still key to avoid unintended consequences. To qualify this result, three considerations can be made. First, the absence of an effect compared to the Control could be due to mistakes reaching a natural lower bound which makes the effect very small and undetectable (due only to Hawthorne-type effects, of which we have shown some evidence). Second, even if the Bottom-up treatment's effect is neutral in terms of performance, it might bring positive effects on workers' well-being, which adds an important motivation to implement it (in addition to mitigate its unintended consequences on performance). Informal reports by the management after the end of the experiment suggested that this was the case. However, in this context it was impossible to measure workers' well-being and satisfaction. Some early studies showed that autonomous motivation has an advantage in terms of providing job satisfaction and well-being (Breaugh 1985; Karasek 1990; Matteson and Ivancevich 1987; Sherman 1989). In a companion, laboratory-based experiment with a data-entry job, Marandola and Savorelli (2021) show that subjects in the participation treatment enjoy the work more, independently of their performance and remuneration scheme. We consider this an important avenue for future research.

Third, if the Bottom-up treatment has a neutral impact on quality provision, the monitoring and recognition system comes at a small cost for the firm, while the cost of top-down rules is potentially high. In addition, as the participative procedure was welcome both by the management and union's representatives, it is good candidate as a parsimonious Pareto-improving practice.

A point we would like to stress is related to the external validity of this experiment. As mentioned, any policy recommendation outside the care homes framework should be cautionary, especially as this experiment shows that (apparently) secondary details might have a large impact on performance in the workplace. Although we could run the controlled experiment for three months, which is a standard testing time for new policies in the firm, studying whether the Bottom-up and Top-down treatment effects would be persistent over longer horizons would be an interesting development of this research.

Our experiment focused on non-monetary incentives, and an obvious question is how these results would extend to the impact of participation in the context of monetary incen-

tives. A few papers (e.g. Marandola and Savorelli 2021, Dal Bó et al. 2019, Mellizo et al. 2014)) provide some laboratory experimental evidence on this topic, but we believe that this will be fruitful ground for future field research.

Appendix A. Tables

Table A.8: Types of mistakes

Users	Organization	Colleagues
1. tone of voice	1. plans/job execution	1. inappropriate mobile use
2. dentures/hearing aids	2. procedures execution	2. public quarrel
3. inattention to communication	3. disposition execution	3. eating from food trays
4. posture	4. shift change	4. other / colleagues
5. aesthetics	5. communication	
6. nails	6. handover process	
7. beard	7. handover reading	
8. hygiene	8. slow handover	
9. clothing	9. other / handover	
10. clothing preparation	10. forms drawing up	
11. clothing control	11. absence at meeting	
12. decorum	12. linen storage	
13. deputise to be faster	13. incontinence storage	
14. handling	14. trash storage	
15. nutrition	15. carts	
16. incontinence aid	16. lateness	
17. wheelchairs	17. uniform	
18. meal tables	18. Personal Protective Equipment (PPE)	
19. wrong information sharing	19. Collective Protective Equipment (CPE)	
20. relationship with relatives	20. break - excessive	
21. room tidiness	21. break - location	
22. gossiping	22. other / organization	
23. use of Italian language		
24. other / end-user		

Table A.9: Poisson regressions. The effect of participation: Bottom-up vs Top-down

	(1)	(2)	(3)	(4)	(5)
Bottom-up	-1.111^{****}	-1.105^{****}	-1.105^{****}	-1.105^{***}	-1.105^{****}
	(0.264)	(0.254)	(0.247)	(0.345)	(0.241)
Persistence	-1.020^{****}	-0.998^{****}	-1.010^{****}	-0.998^{***}	-1.010^{****}
	(0.258)	(0.253)	(0.245)	(0.379)	(0.232)
Working hours		0.028^{****}	0.028^{****}	0.028^{****}	0.028^{****}
		(0.005)	(0.005)	(0.005)	(0.006)
Employment (years)		0.001	-0.021	0.001	-0.021
		(0.014)	(0.013)	(0.012)	(0.016)
Schooling (years)		-0.052^{**}	-0.058^{***}	-0.052^{***}	-0.058^{*}
		(0.022)	(0.022)	(0.020)	(0.031)
Native		0.059	0.009	0.059	0.009
		(0.109)	(0.105)	(0.111)	(0.147)
Female		-0.020	-0.082	-0.020	-0.082
		(0.235)	(0.232)	(0.191)	(0.293)
Age (years)		-0.007	-0.003	-0.007	-0.003
		(0.005)	(0.005)	(0.004)	(0.007)
Hourly pay		0.103	0.235^{*}	0.103	0.235
		(0.137)	(0.134)	(0.108)	(0.155)
Pseudo R^2	0.058	0.078	0.099	0.078	0.099
Fixed effects	NO	NO	YES	NO	YES
Clustered SE	NO	NO	NO	YES	YES
Observations	807	807	807	807	807

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Notes. Dependent Variable: Log(Weekly individual mistakes+1); Poisson estimates (1) with robust standard errors, without controls; (2) with robust standard errors; (3) with robust standard errors and unit-level fixed effects; (4) with cluster-robust standard errors at unit-week level; (5) with cluster-robust standard errors at individual level and unit-level fixed effects. Clustered standard errors are computed over 9*12 clusters at unit-week level, and 73 clusters at individual level.*Bottom-up* is an interaction dummy between time dummy for the treatment period (not shown in the table) and the Bottom-up treatment dummy. *Persistence* is an interaction dummy between time dummy for the post-treatment period (not shown in the table) and the Bottom-up treatment dummy. *Working hours*, effective weekly working hours. Time-invariant controls: *Schooling (years)*, number of years of formal education, excluding professional training. *Employment (years)*, number of years employed in the job. *Hourly pay* is the hourly pay before taxes, in Euro. *Age* is expressed in years. *Female* is a dummy = 1 if the employee is female. *Native* is a dummy = 1 if the employee was not born or not of Italian ancestry. Data include weeks 3-14.

References

Alexander, A., De Smet, A. and Weiss, L. (2020). Decision making in uncertain times, *McKinsey Quarterly [2020-04-24] Available at: http://www. mckinsey. com/business-functions/organization/our-insights/decision-making-in-uncertain-times* .

Argarwal, D., Bersin, J. and Lahiri, G. (2018). The rise of the social enterprise, *Human Capital Trend Report, Deloitte Development LLC* **60**.

Ashraf, N., Bandiera, O. and Lee, S. S. (2014). Awards unbundled: Evidence from a natural field experiment, *Journal of Economic Behavior & Organization* **100**: 44–63.

Attari, S. Z., Gowrisankaran, G., Simpson, T. and Marx, S. M. (2015). Does information feedback from in-home devices reduce electricity use? evidence from a field experiment, *Technical report*, National Bureau of Economic Research.

Bandiera, O., Barankay, I. and Rasul, I. (2009). Social connections and incentives in the workplace: Evidence from personnel data, *Econometrica* **77**(4): 1047–1094.

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

Bartling, B., Fehr, E. and Herz, H. (2014). The intrinsic value of decision rights, *Econometrica* **82**(6): 2005–2039.

Bradler, C., Dur, R., Neckermann, S. and Non, A. (2016). Employee recognition and performance: A field experiment, *Management Science* **62**(11): 3085–3099.

Breaugh, J. A. (1985). The measurement of work autonomy, *Human relations* **38**(6): 551–570.

Cassar, L. and Meier, S. (2018). Nonmonetary incentives and the implications of work as a source of meaning, *Journal of Economic Perspectives* **32**(3): 215–38.

Coch, L. and French, J. R. (1948). Overcoming resistance to change, *Human relations* **1**(4): 512–532.

Cockburn, W. (2020). Covid-19: back to the workplace-adapting workplaces and protecting workers.

Dal Bó, P., Foster, A. and Kamei, K. (2019). The democracy effect: a weights-based identification strategy, *NBER Working Paper no.25724* .

Dal Bó, P., Foster, A. and Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy, *American Economic Review* **100**(5): 2205–29.

Deci, E. L., Koestner, R. and Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation., *Psychological bulletin* **125**(6): 627.

Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior, *Psychological inquiry* **11**(4): 227–268.

Deckop, J. R. and Cirka, C. C. (2000). The risk and reward of a double-edgedsword: Effects of a merit payprogram on intrinsic motivation, *Nonprofit and Voluntary Sector Quarterly* **29**(3): 400–418.

Ellingsen, T. and Johannesson, M. (2007). Paying respect, *Journal of Economic Perspectives* **21**(4): 135–150.

Fehr, E., Herz, H. and Wilkening, T. (2013). The lure of authority: Motivation and incentive effects of power, *American Economic Review* **103**(4): 1325–59.

Folger, R. (1977). Distributive and procedural justice: Combined impact of voice and improvement on experienced inequity., *Journal of personality and social psychology* **35**(2): 108.

Frey, B. S., Benz, M. and Stutzer, A. (2004). Introducing procedural utility: Not only what, but also how matters, *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft* pp. 377–401.

Gagné, M. and Deci, E. L. (2005). Self-determination theory and work motivation, *Journal of Organizational behavior* **26**(4): 331–362.

Gneezy, U. and Rustichini, A. (2000). A fine is a price, *The Journal of Legal Studies* **29**(1): 1–17.

Grolnick, W. S. and Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation., *Journal of Personality and Social Psychology* **52**(5): 890.

Hong, F., Hossain, T., List, J. A. and Tanaka, M. (2018). Testing the theory of multitasking: Evidence from a natural field experiment in chinese factories, *International Economic Review* **59**(2): 511–536.

Jäger, S., Noy, S. and Schoefer, B. (2021). Codetermination and power in the workplace, *Technical report*, Working Paper. Accessed at http://economics. mit. edu/files/21791.

Karasek, R. (1990). *Healthy work: stress, productivity, and the reconstruction of working life*, New York: Basic books.

Kolstad, J. T. (2013). Information and quality when motivation is intrinsic: Evidence from surgeon report cards, *American Economic Review* **103**(7): 2875–2910.

Kube, S., Maréchal, M. A. and Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace, *American Economic Review* **102**(4): 1644–62.

Lawler, E. E. and Hackman, J. R. (1969). Impact of employee participation in the development of pay incentive plans: A field experiment., *Journal of Applied Psychology* **53**(6): 467.

Leonard, K. L. (2008). Is patient satisfaction sensitive to changes in the quality of care? an exploitation of the hawthorne effect, *Journal of Health Economics* **27**(2): 444–459.

Leonard, K. L. and Masatu, M. C. (2010). Using the hawthorne effect to examine the gap between a doctor's best possible practice and actual performance, *Journal of Development Economics* **93**(2): 226–234.

Leonard, K. and Masatu, M. C. (2006). Outpatient process quality evaluation and the hawthorne effect, *Social science & medicine* **63**(9): 2330–2340.

Levitt, S. D. and List, J. A. (2011). Was there really a hawthorne effect at the hawthorne plant? an analysis of the original illumination experiments, *American Economic Journal: Applied Economics* **3**(1): 224–38.

Lind, E. A. and Tyler, T. R. (1988). *The social psychology of procedural justice*, Springer Science & Business Media.

Marandola, G. and Savorelli, L. (2021). Contingent pay, participation, and performance, *MIMEO* .

Matteson, M. T. and Ivancevich, J. M. (1987). *Controlling work stress: Effective human resource and management strategies.*, Jossey-Bass.

Mellizo, P., Carpenter, J. and Matthews, P. H. (2014). Workplace democracy in the lab, *Industrial Relations Journal* **45**(4): 313–328.

Park, R., Blasi, J. R. and Kruse, D. L. (2010). *Shared Capitalism in the US Economy: Prevalence, Characteristics, and Employee Views of Financial Participation in Enterprises*, University of Chicago Press.

Scheflen, K. C., Lawler, E. E. and Hackman, J. R. (1971). Long-term impact of employee participation in the development of pay incentive plans: A field experiment revisited., *Journal of Applied Psychology* **55**(3): 182.

Sheldon, K. M. and Elliot, A. J. (1998). Not all personal goals are personal: Comparing autonomous and controlled reasons for goals as predictors of effort and attainment, *Personality and Social Psychology Bulletin* **24**(5): 546–557.

Sherman, J. D. (1989). Technical supervision and turnover among engineers and technicians: influencing factors in the work environment, *Group & Organization Studies* **14**(4): 411–421.

Thibaut, J. W. and Walker, L. (1975). *Procedural justice: A psychological analysis*, L. Erlbaum Associates.

Van Wanrooy, B., Bewley, H., Bryson, A., Forth, J., Freeth, S., Stokes, L. and Wood, S. (2013). *Employment relations in the shadow of recession: Findings from the 2011 workplace employment relations study*, Macmillan International Higher Education.