ORIGINAL PAPER

Effects of incentive framing on performance and effort: evidence from a medically framed experiment

Mylène Lagarde[1] · Duane Blaauw[2]

Abstract
We study the effects on performance of incentives framed as gains or losses, as well as the effort channels through which individuals increase performance. We also explore potential spill-over effects on a non-incentivised activity. Subjects participated in a medically framed real-effort task under one of the three contracts, varying the type of performance incentive received: (1) no incentive; (2) incentive framed as a gain; or (3) incentive framed as a loss. We find that performance improved similarly with incentives framed as losses or gains. However, individuals increase performance differently under the two frames: potential losses increase participants' performance through a greater attention (fewer mistakes), while bonuses increase the time spent on the rewarded activity. There is no spill-over effect, either negative or positive, on the non-incentivised activity. We discuss the meaning and implications of our results for the design of performance contracts.

Keywords Penalties · Rewards · Laboratory experiment · Prosocial motivation · Intrinsic motivation

JEL Classifications C91 · D64 · I11

1 Introduction

Contracts that link remuneration to the achievement of performance targets are widely used to align the interests of employers and workers in both the private sector (Lazear, 2000) and the public sector (Burgess & Ratto, 2003). The considerable economic literature on the effects of performance contracts acknowledges that incentives improve outcomes through workers' increased effort, but generally

✉ Mylène Lagarde
  M.Lagarde@lse.ac.uk

1 Department of Health Policy, London School of Economics, Houghton Street, London, UK

2 Centre for Health Policy, University of Witwatersrand, Johannesburg, South Africa

⚗ Springer

falls short of unpacking what dimensions of effort change. According to psychologists, incentives increase performance through three potential pathways or changes in effort (Kanfer, 1990). First, incentives can impact the *direction* of effort, which is the choice individuals make to focus on one task or another. Second, incentives can change the *intensity* of effort, or the extent to which workers apply their cognitive resources (i.e. attention or focus exerted to minimise mistakes or increase efficiency, or both). Finally, incentives may affect the *persistence* of effort, which is the time workers spend on a given task. Understanding the mechanisms through which performance is achieved may be particularly important in multi-tasking settings, especially if work is constrained by time limits. If higher performance is achieved through a change in the direction or persistence of effort i.e. individuals engage more in the incentivised activity or spend more time on it, performance in non-incentivised activities is likely to decline due to the automatic reduction in the time available (Holmstrom & Milgrom, 1991). However, if incentives change attentional processes (i.e. the intensity of effort), Kahneman (1973)'s work on attention research suggests that this could increase the overall attentional resource pool available to workers, leading to positive spill-overs on a non-incentivised task (Yechiam & Hochman, 2013b).

In this paper, we explore the direct and indirect effects of incentives on performance and effort channels in a real effort task, under two types of contracts: one that rewards workers' good performance and another that penalises them for poor performance. Prospect theory suggests that because of loss aversion, framing incentives as losses can be more motivating than equivalent rewards (Kahneman & Tversky, 1979). A number of studies have tested this prediction, with mixed results,[1] but very few have explored the mechanisms through which increased performance is achieved under both frames. We contribute to the experimental literature using real effort tasks to study the effects of financial incentives on quantity and quality of output.[2] We also build on studies from cognitive psychology that have shown how potential losses, unlike rewards, heighten the level of attention of subjects (Yechiam & Hochman, 2013c), which can result in increased performance for a non-incentivised task (Yechiam & Hochman, 2013b). We measure the effects of the two contract frames on performance and two possible effort channels (persistence and intensity of effort) in a real-effort experiment that mimics the healthcare context, where performance contracts are ubiquitous and often incomplete.

---

[1] For example, some laboratory (Armantier and Boly 2015; Church et al., 2008; Hannan et al., 2005; Imas et al., 2017) and field experiments (Fryer et al., 2012; Hong et al., 2015; Hossain and List 2012) have found that individuals incentivised with contracts framed as losses perform better than those offered equivalent rewards. Meanwhile, other has failed to detect a difference between the two frames (de Quidt et al., 2017; DellaVigna and Pope 2018; Grolleau et al., 2016).

[2] A number of studies have considered the effects of different types of incentives for quality (Bracha and Fershtman 2013; Carpenter et al., 2010; Eckartz et al., 2012; Hammermann and Mohnen 2014; Rubin et al., 2018; Shurchkov 2012), while others have studied the effects of incentivising quantity on quality (Al-Ubaydli et al., 2015; Fest et al., 2019; Green, 2014; Greiner et al., 2011; Tonin and Vlassopoulos 2015). A few scholars have explored the relative effects and complementarity of incentives for quality and quantity (Kachelmeier et al. 2008; Laske and Schröder 2017).

We describe our experiment design in Sect. 2. Subjects performed a real-effort medically framed task that involved two activities: a routine activity (medical data entry) and a cognitive activity (diagnosis). We randomly allocated participants to a control, gain or loss contract. All contracts included a base pay. In the gain contract, subjects could earn an additional bonus whose size depended on performance in the data entry activity. In the penalty contract, the base pay would be reduced by an amount conditional on their performance in the data entry activity.

We report our results in Sect. 3. Performance in the data entry activity improves in a similar way under the gain and loss contracts, but this is achieved through different behavioural responses. Subjects facing potential losses improve their performance through increased intensity of effort (i.e. reducing the number of errors made), while subjects facing rewards increase persistence of effort (i.e. increased time spent on the incentivised activity). There is no evidence of either negative or positive spill-over effect of either contract on the non-incentivised activity. We discuss these results and their implications in Sect. 4.

2 Experimental setup

2.1 The medical task

We developed a novel real-effort task to mimic the key features of a medical consultation. Before the start of a 10-min period of work,[3] participants receive 10 files of hypothetical patients. A patient file is a laboratory test report that includes 22 two- or three-digit numbers corresponding to standard blood tests.[4] Subjects are then asked to 'manage' as many patients as they can during the 10-min period. Managing a patient is done in three successive steps, with subjects required to validate one step to go to the next one:

(1) Registration: entering patient identifier on the computer interface;
(2) Data entry: entering individual blood test results into a computer mask;
(3) Diagnosis: interpreting the haematology results by identifying the correct pathology from a list of 13 possible diagnoses.[5]

---

[3] The short duration of the task replicates the time constraints under which providers operate in the real world, which can trigger trade-offs between the number of patients seen and the quality of care provided to each patient.

[4] Full Blood Count, Urea & Electrolytes and a Liver Function Test. For more details, see online Appendix C.

[5] All 10 cases present ten unique diagnoses, some of which were arguably easier to identify than others. Harder cases typically corresponded to less common ailments, not necessarily ailments where multiple blood test results were pathological—see online Appendix C for more details.

Although the registration phase is necessary, it is the data entry and diagnosis activities that are the focus of the medical task.[6]

The task design shapes the production process and limits strategic behaviour in several ways. First, each patient managed involved the same 3-step process, meaning that the data entry and diagnosis activities are sequenced (the diagnosis choice automatically follows the end of the data entry activity). Although participants can still choose to devote little time to either activity, they cannot completely ignore one activity if they choose to engage in the other.[7] Second, cherry-picking of easier diagnoses or easier blood test results is possible, but unlikely. Time constraints make identifying easier diagnoses inefficient. Regarding data entry, all reports were qualitatively roughly similar in terms of data entry overall difficulty (i.e. same number of digits). Moreover, while a rational individual might choose to enter easier data entries (e.g. those requiring fewer characters), skipping specific results would require an attention probably more costly than the expected gain.[8]

The welfare benefits of health services for patients are a key feature of healthcare markets, not least because they can form part of providers' utility functions (Arrow, 1963). To incorporate this factor into the task, we followed other experiments conducted in health (Hennig-Schmidt et al., 2011; Lagarde & Blaauw, 2017) and linked subjects' performance to donations to a healthcare charity. Both types of activities (mundane process activities and cognitive ones) are important to achieving high quality of care in the real world. Therefore, they both generate social benefits (i.e. donations to charities) in the experiment. The social incentive was set at R0.20 (USD0.02) for each correctly entered test item and R1.50 (USD0.15) for each correctly identified diagnosis.[9]

2.2 Experimental design

Because doctors' cognitive effort is difficult to observe, performance contracts in health usually focus on routine activities which contribute to better health outcomes.

---

[6] These two activities were chosen to reflect real-world consultations with a patient, during which providers carry out mundane and repetitive activities (taking the patient's vital signs, entering information in the patient's record), but also have to exert some cognitive effort to determine how to manage the patient (i.e. identify the likely diagnosis and most appropriate treatment based on the information available).

[7] The forced sequencing of the two activities was designed to increase the realism of the task: after having questioned and examined a patient, a doctor cannot move on to examining the next one. Instead, she has to decide on the most likely diagnosis (and appropriate treatment). Beyond the medical setting, there are many examples of workers required to undertake different tasks sequentially, e.g. painters prepare a surface then paint it; copy-editors proof-read then format a document; judges study the facts of a crime then choose a sentence; etc.

[8] We found no evidence that participants had a systematic bias in favour of attempting 'easier' blood results rather than difficult ones (see online Appendix G).

[9] In other words, a perfectly managed case created patient benefits worth R5.90 (USD0.59), 75% of which came from the routine activity and 25% from the cognitive activity. This split partly reflects the notion that actual benefits are derived from procedural quality of care, which involves many routine activities. Further, patients are likely to value these observable efforts of providers more than the cognitive effort that they cannot observe or evaluate perfectly.

For example, payments are linked to process measures of quality of care, such as undertaking routine checks or monitoring.[10] Following this logic, we test the effects of contracts that reward performance in the data entry activity, while the diagnosis activity is not incentivised.

Subjects were randomly assigned to one of the three treatments: control, gain or loss treatment.[11] The gain and loss treatments were isomorphic and only differed in their framing. In the gain treatment, subjects earned a base pay of R90, plus an additional bonus worth R10, R20, R30, R40 or R50, depending on the total number of correct entries made.[12] In the loss treatment, the payment specified a base pay of R140, minus a penalty of R10, R20, R30, R40 or R50, if a minimal number of correct entries was not made.[13] In the control treatment, participants received a fixed pay of R105. We used results from a pre-test[14] to calibrate this amount, seeking to equalise the expected remuneration across treatments to control for the income effect on performance.

Participants took part in two consecutive periods of work of 10 min each. Within each group, the second period of work was identical to the first period for half of participants, while the other half was randomised to receiving a reward for each correct diagnosis to stimulate performance in the diagnosis task. Since our objective here is to explore the relative effect of the gain and loss framing of incentives, we focus on the first period, where the control group receives no incentive, and use the second one only as a robustness check.[15]

The experiment was run using the software z-Tree (Fischbacher, 2007)—see online Appendix C for screenshots. Participants received an attendance fee of R30 (USD2.90) and were allocated to a workstation according to a random blocked design to obtain an equal number of participants per treatment (see online Appendix D for more details on experimental procedures). Specific instructions on the computer screen explained how they would be remunerated in the task (see online Appendix E). Before the task started, subjects were informed that patient benefits generated in the task would translate into actual donations for healthcare delivery and they could select their preferred charity from a list of five.[16] At the end of the

---

[10] For example, in the management of patients with chronic conditions, performance contracts typically reward the monitoring of blood pressure, blood glucose, and other risk factors at regular intervals.

[11] Note that the experiment initially included a fourth treatment, which has been excluded due to flaws in its design and implementation. This exclusion does not compromise the internal validity of the rest of the experiment presented here. Online Appendix B provides more details.

[12] If subjects entered between 100–109, 110–119, 120–129, 130–39 correct numbers or 140 or more correct numbers, they earned an additional bonus of R10, R20, R30, R40, and R50, respectively.

[13] If subjects entered fewer than 100 correct numbers, between 100–09, 110–19, 120–29 or 130–39 numbers, they lost R50, R40, R30, R20 or R10, respectively.

[14] The pre-test was organised with 14 students from the same subject pool, who were working under the conditions of the gain treatment described in the text.

[15] We can focus on period 1 without threatening the internal validity of the experiment because participants were informed they would be paid for one of the two periods chosen at random, and there was no anticipation effect since they did not know at the start of period 1 what incentives they would face in period 2.

[16] Subjects could choose from the following five local health charities: Witwatersrand Hospice; SOS

session, each participant received their payment anonymously after completing a short questionnaire capturing basic socio-demographic information.

A total of 180 medical students participated in 11 experimental sessions. A session lasted approximately 45 minutes and on average participants earned R118.3 (USD11.4) in addition to the attendance fee, and a total of R5,223.9 (USD505) was transferred to charities. Participants were fifth year medical students[17] from the Medical School at the University of Witwatersrand in Johannesburg, South Africa. Their characteristics were similar across all treatment groups (see Table A1 in online Appendix A).[18]

2.3 Testable hypotheses and data

We formulate the following five testable hypotheses:

H1: following standard economic theory, financial incentives in the loss and gain treatments will lead to higher performance in data entry.

H2: loss aversion theory (Kahneman & Tversky, 1979) predicts that performance in data entry will be higher in the loss treatment compared to the gain treatment.

H3: according to studies in psychology, the mechanism behind increased performance with incentives framed as losses is not the higher subjective weight given to losses compared to gains (Kahneman & Tversky, 1979), but the fact that losses create a physiological arousal in subjects which draw their attention to the task more than gains (Yechiam & Hochman, 2013a, c). Hence, in our context, this "loss attention" model predicts that higher performance in data entry will be achieved through increased attentional investment (i.e. higher accuracy).

H4: the theory of incomplete contracts (Holmstrom & Milgrom, 1991) predicts a reduction in effort (time) invested in the diagnosis activity when data entry is incentivised, leading to a reduction in performance. However, in our setting, subjects (medical students) might be intrinsically motivated to perform in this task, hence limiting the negative effect of incentives on the non-incentivised activity.

H5: according to attention research (Kahneman 1973), when individuals work on several tasks, an increase in attention in one task can occur through two different mechanisms: (i) a change in the relative allocation of attention from one task to the other or (ii) an overall increase in attentional resources, which will benefit all tasks proportionally to the initial allocation of resources. In line with the prediction of the "loss attention" model (H3), losses are expected to increase the overall attentional

---

Footnote 16 (continued)

Children's Villages; South African National Tuberculosis Association; Cancer Association of South Africa; or Thusanani Children's Foundation.

[17] These students are all in the clinical training phase of their medical education, undertaking clinical rotations and possessing sufficient clinical knowledge to complete the diagnosis activity.

[18] In particular, participants did not differ in their academic performance (based on exam results from the previous year), their ability to interpret blood test results (based on a knowledge test of seven clinical conditions used in the experiment) or their personality traits (measured with an abridged version of the Big 5 scale (Thompson 2008)).

resources, leading to a positive spill-over effect on the diagnosis task even if it is not directly incentivised (Yechiam & Hochman, 2013b).

Performance in the data entry activity is measured by the total number of correct entries made over the period, since this is the performance target in the loss and gain treatments. Similarly, we consider the number of correct diagnoses made as the performance measure in the diagnosis activity.

To explore how incentives affect two possible channels of effort,[19] we first measure the persistence of effort as the total time spent on an activity.[20] Second, in the absence of a physiological measure of attention, we use accuracy in an activity (proportion of correct attempts out of total attempts made) as a proxy for effort intensity, assuming that increased attention reduces errors. However, if individuals seek to minimise errors (improve performance) by double-checking *ex-post* that their response is correct, the two channels of effort may not be independent from each other, as verifying one's responses takes time. In online Appendix F, we show that the correlation between our measures of effort intensity and persistence is low, thus providing support to the notion that performance is increased by being more focussed and avoiding mistakes *ex-ante*.

Table A2 in online Appendix A provides descriptive statistics of all performance measures for the three groups.

3 Results

3.1 Effect of framing on performance in the incentivised activity

We first explore the effects of incentives on the targeted activity (data entry). Evidence from the distribution of performance results (Fig. 1) supports hypothesis (*H1*) that performance in data entry is higher in the two incentive treatments. Overall, about 20% more correct entries are made under performance contracts: compared to the 96.9 entries in the control treatment, participants made 117.8 correct entries in the loss treatment ($p = 0.008$ two-sided Mann–Whitney U-Test, hereafter MW test), and 116.6 in the gain treatment ($p = 0.023$ MW test). However, there is no evidence supporting the prediction (*H2*) of loss aversion theory that performance under the loss contract is higher than under the gain contract ($p = 0.839$, MW test). This result is robust to the inclusion of subjects' characteristics in a regression (Table 1, column 2), and we fail to reject the null hypothesis that individuals' performance is the same under both frames ($p = 0.648$, test of equality of coefficients).

---

[19] In our setup, participants cannot choose the direction of their effort since they automatically face both activities. However they can decide to ignore one activity by spending no time on it—this is captured by the persistence of effort.

[20] Note that all three stages described in Sect. 2.1 are time-stamped separately—hence the overall time spent on data entry and diagnosis does not sum to 600 s: the first stage (registration_) accounts for the 'missing' time.

3.2 Effect of framing on the channels of effort

Looking at the effort channels through which higher performance was achieved under the two contracts, the raw data (see Table A2 and Figure A1 in online Appendix A) show that subjects in the gain contract achieve high performance by spending more time on data entry compared to the other two groups. On average, they spend 30 more seconds on this activity than subjects in the control group ($p = 0.019$), and 25 more second than those in the loss group ($p = 0.012$). By contrast, there is no difference between the control and loss group ($p=0.995$). Regression results presented in Table 2 confirm these findings, which remain robust to the inclusion of individual controls (Column 2).

Next we consider the intensity of effort, proxied by accuracy in the task. Participants in the loss group achieve near perfect accuracy: with 97.8% of entries correctly made, this is 9 percentage points (pp) higher than in the control group ($p = 0.014$) and 5.2 pp higher than the gain treatment ($p = 0.042$).[21] This level of attention and limited number of mistakes are consistent with the notion of heightened attention dedicated to the task due to the threat of losses (H3). The results are robust to controlling for additional demographic characteristics (Table 2, column 4).

3.3 Effects on the non-incentivised activity

Next, we consider the effects of the contract frames on the non-incentivised diagnosis activity. Unlike what is predicted by standard economic theory (*H4*), performance (number of correct diagnoses) does not decrease under either incomplete contract. Subjects in the control group make 3.2 correct diagnoses, against 3.5 correct diagnoses ($p=0.189$) in the gain treatment, and 3.3 correct diagnoses in the loss treatment ($p=0.276$). This null result is robust to the inclusion of individual characteristics (Table 3, column 2). Overall, this result supports the notion that individuals are intrinsically motivated to perform well in this task.

Turning to the channels of effort, we fail to detect any difference in intensity of effort (accuracy) between the incentives and the control group (Table 3, columns 3 and 4), or across incentive frames. In the loss treatment, there is no evidence of a positive spill-over effect of the heightened attention in data entry on the diagnosis activity (H5). Consistent with the result that they spent more time on data entry, subjects under the gain contract spent nearly 23 fewer seconds on the diagnosis activity, while there is no evidence that under the loss framed incentives, subjects spent less time on diagnosis (Table 3, columns 5 and 6).

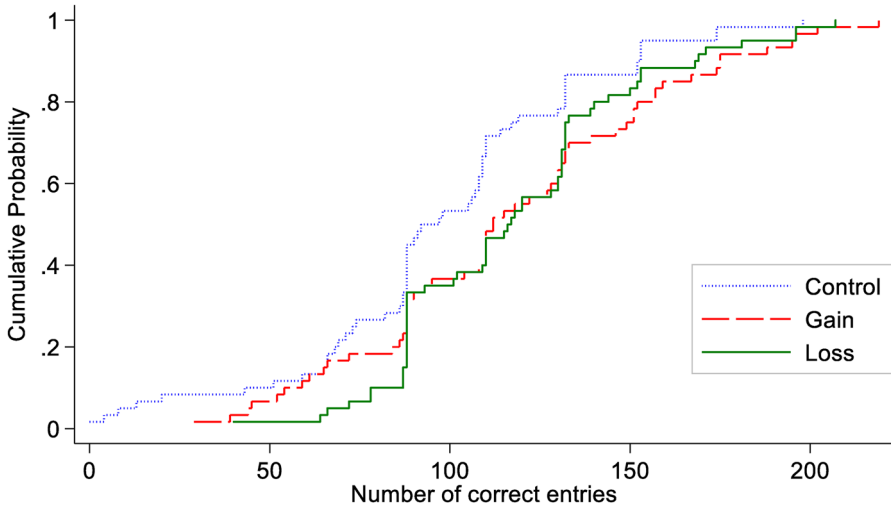[21] By contrast, accuracy under the gain framing is not different from that in the control group ($p=0.367$).

Fig. 1 Performance in the data entry activity, by treatment

Table 1 Impact of financial incentives on performance in data entry

	Performance in data entry	
	(1)	(2)
Gain	19.683*** (7.360)	16.127* (6.771)
Loss	20.900*** (7.360)	19.213*** (6.735)
Individual controls	No	Yes
Mean in control group (SD)	96.93 (41.24)	
Observations	180	180
R^2	0.054	0.271

This table reports the OLS regression of performance measures in the data entry activity on dummy variables of the treatment conditions. The dependent variable is equal to the total number of correct entries made by the participant. Individual controls include age, gender, ethnicity, grade obtained the previous year, knowledge of blood test result interpretation and personality traits (Big 5 inventory). Robust standard errors are in parentheses

***$p < .001$ **$p < .01$, *$p < .05$

3.4 Choice of optimal effort allocation

Even though a correct diagnosis generated no private monetary gains, our data show that participants spent on average more than 20% percent of their time on diagnosis,[22] which is inconsistent with purely selfish financial motives. Beyond intrinsic

---

[22] Participants in the gain treatment are those who spent the least amount of time on this activity: on average 122 s, or 20.3 percent of their working time.

Table 2 Impact of financial incentives on effort persistence and intensity in data entry

Dependent variable	Effort persistence (time spent)		Effort intensity (accuracy)	
	(1)	(2)	(3)	(4)
Gain	29.807*** (10.901)	29.053* (11.164)	0.040 (0.036)	0.035 (0.037)
Loss	4.465 (10.901)	5.408 (11.106)	0.092* (0.036)	0.084* (0.037)
Individual controls	No	Yes	No	Yes
Mean in control group (SD)	390.01 (69.80)		0.89 (0.28)	
Observations	180	180	180	180
R^2	0.047	0.090	0.035	0.073

This table reports the OLS regression of effort measures in the data entry activity on dummy variables of the treatment conditions. In columns 1–2, the dependent variable is defined as the time spent on data entry by an individual over the entire period of work (in seconds). In columns 3–4, the dependent variable is the subject's accuracy in the task, calculated as the ratio of total number of correct entries over all attempted entries during the period. Individual controls include age, gender, ethnicity, grade obtained the previous year, knowledge of blood test result interpretation and personality traits (Big 5 inventory). Robust standard errors are in parentheses

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

motivation, could altruistic motives explain this behaviour? To answer this question, we consider how altruistic subjects *should* allocate their time between diagnosis and data entry. This choice depends on relative expected (social) earnings per unit of time spent in both activities. Given that social rewards for each activity are fixed, participants should allocate their effort between data entry or diagnosis depending on their relative abilities (a combination of their speed, i.e. average time per output produced, and their accuracy). A rational decision-maker should focus on data entry as long as each second spent on this activity yields higher returns than a second spent on diagnosis. Given the social benefits attached to the two activities, a subject should focus on data entry if the time she needs to obtain a *correct* diagnosis is at least 7.5 times higher than the time per correct data entry.[23] Taking the median abilities in the control group as proxies[24]—i.e. 4.12 seconds per correct data entry, and 47.95 per correct diagnosis (see Table A2)—the optimal strategy is to focus entirely on data entry to maximise charity donations. We find that the median subject dedicating all of her working time to data entry could raise R25.9 for charity, by making just under 130 correct entries, which is higher than the median performance observed in all treatments.[25] Note that there is one case where it would be optimal to have the reverse strategy: focus on diagnoses and spend any remaining time on data

[23] This is because participants receive R0.20 per correct data entry and R1.50 for a correct diagnosis. See online Appendix H for more details on the optimal strategy.

[24] We take the median performance rather than the average because two outliers significantly distort the average values to 6.97 s per correct entry and 52.36 s per correct diagnosis. However, even taking the average values, the optimal strategy is to focus on data entry.

[25] In the absence of measures of ability independent from the treatments, it is impossible to determine whether individuals behave optimally or not, since abilities calculated in the data are endogenous to the time spent on each activity.

Table 3 Impact of financial incentives on effort and performance in diagnosis identification (non-incentivised activity)

Dependent variable	Performance (# of diagnoses correctly identified)		Effort intensity (accuracy)		Effort persistence (time spent)	
	(1)	(2)	(3)	(4)	(5)	(6)
Gain	0.300 (0.294)	0.203 (0.278)	−0.023 (0.050)	−0.022 (0.051)	−22.606** (10.121)	−22.851** (10.442)
Loss	0.167 (0.294)	0.139 (0.276)	−0.004 (0.050)	−0.007 (0.050)	5.197 (10.121)	3.960 (10.388)
Individual controls	No	Yes	No	Yes	No	Yes
Mean in control group (SD)	3.18 (1.53)		0.67 (0.29)		144.65 (62.13)	
Observations	180	180	180	180	180	180
R^2	0.006	0.190	0.002	0.060	0.046	0.075

This table reports the OLS regression of performance measures in the diagnosis identification activity on dummy variables of the treatment conditions. The dependent variable in columns 1–2 is equal to an individual's total number of correct diagnoses identified over the 10-mn period of work. In columns 3–4, the dependent variable (effort intensity) is measured by the subject's accuracy in diagnosis identification, calculated as the ratio of total number of correct diagnoses over all attempted ones during the period. In columns 5–6, the dependent variable is defined as the time spent by an individual on diagnosis identification over the entire period of work (in seconds). Individual controls include age, gender, ethnicity, grade obtained the previous year, knowledge of blood test result interpretation and personality traits (Big 5 inventory). Robust standard errors in parentheses

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

Ⓐ Springer

entry. This would be for participants with high abilities in diagnosis identification (speed and accuracy), but median skills in data entry—see online Appendix H for a description of this scenario.[26] However, this combination of skills is highly unusual; only one individual in the sample fits this profile.

3.5 Robustness check

Results from a second period of 10-min undertaken in the same conditions by half of respondents are shown in Table A3 in online Appendix A. The results confirm the similar positive effects of incentives on performance (Table A3 columns 1–2), with no evidence of a higher performance increase with incentives framed as losses ($p = 0.964$). As in period 1, higher performance in the Gain treatment is achieved through an increase in effort persistence (Table A3 columns 3–4), but no increase in effort intensity (Table A3 columns 5–6). Evidence also confirms the notion that incentives framed as losses (but not gains) increase attention (H3), proxied by accuracy in data entry. However, we also find that individuals in the Loss treatment achieved a higher performance in data entry through an increase in effort persistence. This may not be entirely surprising in a context where 74 percent of subjects under both incentive frames spent more time on data entry in period 2 relative to period 1, possibly after realising that time spent on data entry could lead to higher returns.[27] As a result, in both treatments, there is a substitution of effort persistence away from the diagnosis activity. Yet this does not translate into a significant reduction in performance in diagnosis identification.

4 Discussion and conclusion

Using a novel medically framed real-effort experiment, we evaluate the impact of incomplete contracts, framed as gains or losses, on subject performance and effort channels. We find that both types of incentives lead to increased performance (H1), with no evidence that losses triggered a greater response than gains (H2). This null result echoes those of studies where subjects are told in advance which performance target they have to reach (de Quidt et al., 2017). Another possible explanation lies in the fact that incentives used were relatively small for participants, which has been found to reduce the likelihood of loss aversion (Mukherjee et al., 2017).[28]

---

[26] This case is described in columns B3 and B4 of Table H1 of online Appendix H. Individuals with top abilities in diagnosis and median abilities in data entry maximise social earnings when they prioritise diagnosis (column B4) with R27.94 raised, against R25.93 when prioritising data entry (column B3).

[27] This strategy is indeed superior: all respondents who took part in these two identical periods earned the same or more money in the second period relative to the first one.

[28] In this experiment, although the contingent part of the remuneration was large compared to the fixed component, incentives were relatively small in absolute terms, representing about 10% of a day's worth of work. It is possible that subjects did not value losses differently from gains at such levels of remuneration.

We find that the two framings triggered different behavioural responses. Consistent with the loss attention model (H3), subjects facing losses achieved higher levels of performance by increasing their attention in data entry. The potential implications could be far from trivial in settings where errors are costly. In healthcare, for example, greater attention can reduce medical errors and adverse patient outcomes (Yang et al., 2018). However, generalising beyond the lab is challenging, especially from a 10-min period of work. Sustaining increased attention over longer periods of time could become costly and generate other trade-offs, as seen here in the second period of work. More research will be needed to explore the context in which performance contracts framed as losses can be used to reap the benefits of increased effort intensity.

How to explain that performance in the diagnosis activity did not fall when data entry was incentivised (H4)? Several factors may explain this result. First, performance in that task may be less sensitive to effort exerted and more to the (random) difficulty of cases seen. As mentioned in the description of the task, cases included easy and common diagnoses—needing little reflection—and more uncommon ailments, requiring more advanced knowledge, so that many medical students could have failed to identify the latter, even when spending a lot of time on them. The imperfect relationship between time and performance could partly explain why, even when less time was spent on diagnoses, performance still did not fall. Second, in both treatments, this result may be driven by the interaction of the task design (sequencing of the two activities) with the subjects' intrinsic motivation. Since the diagnosis screen automatically appeared when subjects finished a data entry sheet, not only was it impossible to skip the diagnosis activity entirely, but it was also a reminder of the potential satisfaction derived from doing this task, which was deemed interesting by most[29] and echoed participants' identity (Akerlof & Kranton, 2000).[30] Theoretical models suggest that if agents are intrinsically motivated to exert effort (Besley & Ghatak, 2018), the expected adverse effects of incomplete contracts may be limited, as it has been found in empirical studies in health.[31] In the loss treatment, this result is also likely driven by the fact that increased performance in data entry is obtained through increased intensity of effort (H5), which does not deplete resources available for diagnosis identification.[32] Lastly, even in the absence of monetary reward for data entry, as shown in Section 3.4, a prosocial subject willing to maximise the payoff of the charity should already prioritise this activity over

---

[29] Across all treatments, 82.6% found the diagnosis activity "interesting" and only 2.7% found it "boring". By contrast, only 57.7% found data entry "interesting" and 10% found it "boring".

[30] According to Akerlof and Kranton (2000), individuals would derive utility both from the act of identifying a diagnosis itself, and from the fact that doing this conforms to the identity of the medical doctor they aspire to be.

[31] The few studies looking at non-incentivised activities in performance contracts in health have failed to find evidence of adverse effects (Campbell et al., 2009; Mullen et al., 2010),

[32] This interpretation is supported by suggestive evidence from the second period, mentioned before: when performance in data entry requires both higher attention and more time, we observe a reduction in the number of correct diagnoses found, although not quite significant at conventional levels ($p=0.109$).

diagnosis in the control group. A different result might emerge in a setup where the social incentives to identify correct diagnoses at baseline were higher.

The next question is why did we not observe a positive spill-over effect of heightened attention on diagnosis performance, resulting from an increase in the pool of attentional resources (H5)? The answer may lie, again, in the non-linear relationship between effort and performance: unlike the simple decision task used by Yechiam and Hochman (2013b), identifying a diagnosis correctly not only requires attention, but also knowledge: as several diagnoses were particularly hard to find, the scope for improved performance was limited for most participants.

Our results highlight the importance for employers of considering not simply *if* incentive contracts are effective at increasing the rewarded dimension of performance, but also *how* these contracts improve performance. Indeed, our findings suggest that the behavioural implications of different incentive designs may be far from trivial, especially in settings where individuals undertake different types of tasks. In the setting we simulated, where remuneration was linked to the quantity of output of good quality produced, although incentives framed as gains and losses both achieved a similar result, the loss frame led to the virtual elimination of wastage in production, as participants no longer made errors. This is a key upside, which could have important implications for performance pay in settings where both quantity and quality of outputs are critical. On the other hand, incentives framed as gains reduced some of the effort that workers put in the non-rewarded activity. While this had no further consequence on performance in our experiment, for reasons discussed above, it could become a liability in a context where individuals are not intrinsically motivated. In general, our findings should encourage researchers and policy-makers to explore further the relative effects of incentives framed as losses and gains for incentivising workers, and in particular healthcare providers.

References

Akerlof, G., & Kranton, R. (2000). Economics and identity. *The Quaterly Journal of Economics, 115*(3), 715–753.

Al-Ubaydli, O., Andersen, S., Gneezy, U., & List, J. A. (2015). Carrots that look like sticks: toward an understanding of multitasking incentive schemes. *Southern Economic Journal, 81*(3), 538–561. https://doi.org/10.4284/0038-4038-2013.248

Armantier, O., & Boly, A. (2015). Framing of incentives and effort provision. *International Economic Review, 56*(3), 917–938. https://doi.org/10.1111/iere.12126

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review, 53*(5), 941–943.

Besley, T., & Ghatak, M. (2018). Prosocial motivation and incentives. *Annual Review of Economics, 10*(1), 411–438. https://doi.org/10.1146/annurev-economics-063016-103739

Bracha, A., & Fershtman, C. (2013). Competitive incentives: working harder or working smarter? *Management Science, 59*(4), 771–781.

Burgess, S., & Ratto, M. (2003). The role of incentives in the public sector: issues and evidence. *Oxford Review of Economic Policy, 19*(2), 285–300.

Campbell, S. M., Reeves, D., Kontopantelis, E., Sibbald, B., & Roland, M. (2009). Effects of pay for performance on the quality of primary care in England. *New England Journal of Medicine, 361*(4), 368–378. https://doi.org/10.1056/NEJMsa0807651

Carpenter, J., Matthews, P. H., & Schirm, J. (2010). Tournaments and office politics: evidence from a real effort experiment. *The American Economic Review, 100*(1), 504–517. https://doi.org/10.1257/aer.100.1.504

Church, B. K., Libby, T., & Zhang, P. (2008). Contracting frame and individual behavior: experimental evidence. *Journal of Management Accounting Research, 20*(1), 153–168.

de Quidt, J., Fallucchi, F., Kölle, F., Nosenzo, D., & Quercia, S. (2017). Bonus versus penalty: how robust are the effects of contract framing? *Journal of the Economic Science Association, 3*(2), 174–182. https://doi.org/10.1007/s40881-017-0039-9

DellaVigna, S., & Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies, 85*(2), 1029–1069. https://doi.org/10.1093/restud/rdx033

Eckartz, K., Kirchkamp, O., & Schunk, D. (2012). How do incentives affect creativity? In. CESifo Working Paper Series No. 4049.

Fest, S., Kvaloy, O., Nieken, P., & Schöttner, A. (2019). Motivation and incentives in an online labor market. In. CESifo Working Paper No. 7526: https://ssrn.com/abstract=3343857.

Fischbacher, U. (2007). z-Tree: zurich toolbox for ready-made economic experiments. *Experimental Economics, 10*(2), 171–178.

Fryer, R., Levitt, S., List, J., & Sadoff, S. (2012). Enhancing the Efficacy of Teacher Incentives through loss aversion: a Field Experiment. *NBER Working Paper No. 18237*.

Green, E. P. (2014). Payment systems in the healthcare industry: An experimental study of physician incentives. *Journal of Economic Behavior & Organization, 106*, 367–378. https://doi.org/10.1016/j.jebo.2014.05.009

Greiner, B., Ockenfels, A., & Werner, P. (2011). Wage transparency and performance: a real-effort experiment. *Economics Letters, 111*(3), 236–238. https://doi.org/10.1016/j.econlet.2011.02.015

Grolleau, G., Kocher, M. G., & Sutan, A. (2016). Cheating and loss aversion: do people cheat more to avoid a loss? *Management Science, 62*(12), 3428–3438. https://doi.org/10.1287/mnsc.2015.2313

Hammermann, A., & Mohnen, A. (2014). The pric(z)e of hard work: different incentive effects of non-monetary and monetary prizes. *Journal of Economic Psychology, 43*, 1–15. https://doi.org/10.1016/j.joep.2014.04.003

Hannan, R. L., Hoffman, V. B., & Moser, D. V. (2005). Bonus versus penalty: does contract frame affect employee effort? In A. Rapoport & R. Zwick (Eds.), *Experimental Business Research (Vol* (Vol. II, pp. 151–169). Springer.

Hennig-Schmidt, H., Selten, R., & Wiesen, D. (2011). How payment systems affect physicians' provision behaviour—an experimental investigation. *Journal of Health Economics, 30*(4), 637–646. https://doi.org/10.1016/j.jhealeco.2011.05.001

Holmstrom, B., & Milgrom, P. (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization, 7*, 24–52.

Hong, F., Hossain, T., & List, J. A. (2015). Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization, 118*, 372–382. https://doi.org/10.1016/j.jebo.2015.02.014

Hossain, T., & List, J. A. (2012). The behavioralist visits the factory: increasing productivity using simple framing manipulations. *Management Science, 58*, 2151–2167.

Imas, A., Sadoff, S., & Samek, A. (2017). Do people anticipate loss aversion? *Management Science, 63*(5), 1271–1284. https://doi.org/10.1287/mnsc.2015.2402

Kachelmeier, S. J., Reichert, B. E., & Williamson, M. G. (2008). Measuring and motivating quantity, creativity, or both. *Journal of Accounting Research, 46*(2), 341–373.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall, Englewood Cliffs, NJ.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*(2), 263–291.

Kanfer, R. (1990). Motivation theory and industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*.

Lagarde, M., & Blaauw, D. (2017). Physicians' responses to financial and social incentives: a medically framed real effort experiment. *Social Science and Medicine, 179*, 147–159. https://doi.org/10.1016/j.socscimed.2017.03.002

Laske, K., & Schröder. (2017). Quantity, Quality and Originality: The Effects of Incentives on Creativity. *Annual Conference 2017 (Vienna): Alternative Structures for Money and Banking*. Retrieved from https://ideas.repec.org/p/zbw/vfsc17/168151.html

Lazear, E. (2000). The power of incentives. *American Economic Review, 90*(2), 410–414.

Mukherjee, S., Sahay, A., Pammi, V. S. C., & Srinivasan, N. (2017). Is loss-aversion magnitude-dependent? Measuring prospective affective judgments regarding gains and losses. *Judgment and Decision Making, 12*(1), 81–89.

Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics, 41*(1), 64–91.

Rubin, J., Samek, A., & Sheremeta, R. M. (2018). Loss aversion and the quantity–quality tradeoff. *Experimental Economics, 21*(2), 292–315. https://doi.org/10.1007/s10683-017-9544-1

Shurchkov, O. (2012). Under pressure: gender differences in output quality and quantity under competition and time constraints. *Journal of the European Economic Association, 10*(5), 1189–1213. https://doi.org/10.1111/j.1542-4774.2012.01084.x

Thompson, E. R. (2008). Development and validation of an international english big-five mini-markers. *Personality and Individual Differences, 45*(6), 542–548. https://doi.org/10.1016/j.paid.2008.06.013

Tonin, M., & Vlassopoulos, M. (2015). Corporate philanthropy and productivity: evidence from an online real effort experiment. *Management Science, 61*(8), 1795–1811. https://doi.org/10.1287/mnsc.2014.1985

Yang, H.-C., Islam, M. D. M., & Li, Y.-C. (2018). Monitor, reduce and prevent the adverse outcomes for ensuring patient safety. *International Journal for Quality in Health Care, 30*(6), 415–415. https://doi.org/10.1093/intqhc/mzy109

Yechiam, E., & Hochman, G. (2013a). Loss-aversion or loss-attention: the impact of losses on cognitive performance. *Cognitive Psychology, 66*(2), 212–231. https://doi.org/10.1016/j.cogpsych.2012.12.001

Yechiam, E., & Hochman, G. (2013b). Loss attention in a dual-task setting. *Psychological Science, 25*(2), 494–502. https://doi.org/10.1177/0956797613510725

Yechiam, E., & Hochman, G. (2013c). Losses as modulators of attention: review and analysis of the unique effects of losses over gains. *Psychological Bulletin, 139*(2), 497–518. https://doi.org/10.1037/a0029383