Does Pay Inequality Affect Worker Effort? An Assessment of Existing Laboratory Designs

Marco Fongoni

amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Does Pay Inequality Affect Worker Effort? An Assessment of Existing Laboratory Designs

Marco Fongoni

WP 2022- Nr 30

Does Pay Inequality Affect Worker Effort? An Assessment of Existing Laboratory Designs*

Marco Fongoni†

Aix-Marseille Univ, CNRS, AMSE

November, 2022

Abstract

This paper develops a theoretical framework to think about employees' effort choices, and applies this framework to assess the ability of existing laboratory designs to identify the effect of pay inequality on worker effort. The analysis shows that failure to control for a number of confounds—such as reciprocity towards the employer in multi-lateral gift-exchange games (vertical fairness), or the incentive to increase effort when feeling underpaid under piece rates (income targeting)—may lead to inaccurate interpretation of evidence of treatment effects. In light of these findings, the paper provides a set of recommendations on how to improve identification in the design of laboratory experiments in the future.

Keywords: pay inequality; effort; laboratory experiments; reference dependence; fairness.

JEL classification: C91, D91, J31, J33, M52.

*I would like to thank Werner Bönte, Beatrice Boulu-Reshef, Alex Dickson, Markus Gehrsitz, Jasper Hepp, Nicolas Jacquemet, Jonathan Norris, Sofie Waltl, as well as participants at the Economics and Psychology seminar at Université Paris 1 (Pantheon-Sorbonne), at the Bergische Universität Wuppertal departmental seminar, at the EALE Conference, in Padova, Italy, at the IMERA conference and Ecolunch Seminar in Marseille, France, for valuable discussion and feedback. I acknowledge funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University - A*MIDEX.

†E-mail: marco.fongoni@univ-amu.fr.

1 Introduction

The question of whether pay inequality affects worker effort is as important as it is puzzling. Among compensation managers, pay secrecy and pay compression practices are often justified by a common concern that pay inequality is detrimental for morale and effort (Bewley, 2007). These considerations are also relevant for the development of business cycle models of the labour market that rely on the assumption of internal pay equity concerns to generate wage rigidity and explain unemployment fluctuations (e.g. Gertler and Trigari (2009) Snell and Thomas (2010)). These sorts of concerns also find ample support in the economics and behavioural science literature that suggests social comparisons in the dimension of pay have a major role in influencing employees' perception of fairness and morale in the workplace (see, e.g. Akerlof and Yellen (1990) and Blau (1994)).[1] Nevertheless, to date, it would be presumptuous to conclude that our understanding of the effect of pay inequality on effort is complete.

To investigate this question economists have implemented a number of empirical methodologies, spanning from administrative data analysis (e.g. Smith (2015)) and field surveys (e.g. Bewley (1999)), to laboratory (e.g. Charness and Kuhn (2007)), field (e.g. Cohn, Fehr, Herrmann, and Schneider (2014)) and natural experiments (e.g. Cullen and Perez-Truglia (2022)). With few recent exceptions,[2] the unobservable nature of worker effort and perceived fairness has led to a widespread use of controlled laboratory experiments. However, despite a rich body of literature and ample evidence, it is perhaps surprising that experimental evidence on the effect of pay inequality on effort is mixed, and remains yet inconclusive.[3]

This paper contributes to this literature from the perspective of laboratory experiments. The aim of the paper is to formally investigate the existence of potential confounds that may have impaired the identification of treatment effects in existing designs—perhaps contributing to the evidence being mixed—and, on this basis, to offer a set of recommendations on how to improve identification in the future. For instance, are experiments adopting multi-lateral gift-exchange games capable of disentangling between the morale effect of pay inequality (horizontal

[1] See also Goodman (1974), Frank (1984), Loewenstein and Bazerman (1989), Ordóñez, Connoly, and Coughlan (2000) and McDonald, Nikiforakis, Olekalns, and Sibly (2013).

[2] Such as the natural field experiments of Flynn (2022) and Cullen and Perez-Truglia (2022)

[3] Among laboratory experiments, some studies report no evidence in support of this hypothesis (e.g. Charness and Kuhn (2007), and Bartling and von Siemens (2011)), while other studies report that only disadvantageous pay inequality has a significantly stronger and negative impact on employees' effort (e.g. Gächter and Thöni (2010), Nosenzo (2013), and Bracha, Gneezy, and Loewenstein (2015)). Among field experiments the evidence is also mixed: while Cohn et al. (2014) and Breza, Kaur, and Shamdasani (2018) report some evidence of the effect of disadvantageous pay inequality on effort, Hennig-Schmidt, Sadrieh, and Rockenbach (2010) and Sseruyange and Bulte (2020) find no evidence.

fairness) and the incentive for reciprocity towards the employer (vertical fairness) on employees' choice of effort? Else, are experiments adopting piece-rate payments considering the incentive of underpaid employees to increase their effort in order to increase their final earnings (income targeting) and reduce the expected inequality? In the former case, horizontal and vertical fairness considerations are two concurring factors in explaining workers' effort responses to pay inequality, while in the latter case, income targeting behaviour may dampen, or even offset, any detrimental effect of disadvantageous pay inequality on effort. In both cases, failure to appropriately control for these aspects may lead to inaccurate interpretations of treatment effects.

To investigate these issues the paper develops a model of worker preferences that is suitable to analytically think about the impact of pay inequality, reciprocity and income targeting on a worker's choice of effort. The model is then applied to characterise subjects' effort responses to a variety of pay treatments adopted by the existing literature and, subsequently, to assess whether evidence of treatment effects is in fact determined by the morale effect of pay inequality, or whether there are other factors at play. In so doing, a key contribution of the paper is to advance a novel conceptual framework which fits two purposes: i) to classify and more clearly compare existing designs in terms of treatments and testing approaches and ii) to provide a formal analysis of subjects' effort choices in the context of experiments. The advantage of pursuing this approach is twofold. First, it enables to analyse subjects' effort choices in the context of the existing experimental literature under a single theoretical framework, and to establish more precisely which features of laboratory designs are particularly effective in the identification of the effect of pay inequality on effort. Second, it also enables to uncover which aspects of existing designs may confound this identification, and therefore how to change them to improve the design of experiments in the future.[4] This methodological investigation is important, not only to gain a clearer understanding of the existing evidence, but also to guide the development of further research on this topic. In fact, although field and natural experiments are not discussed in depth, the conclusions and recommendations advanced in this paper could be informative to studies adopting these methods.

The theoretical model used in the assessment draws from the literature on distributional preferences (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000), reciprocity (Rabin, 1993;

[4]The present paper does not deal with design features such as the choice between the direct response method and the strategy method (see Gächter, Nosenzo, and Sefton (2012) for a thorough discussion of this), or the choice between real effort and stated effort (which is addressed by Charness, Gneezy, and Henderson (2018)).

Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006) and reference dependence and loss aversion in the evaluation of pay by workers (Goette, Huffman, and Fehr, 2004; Dickson and Fongoni, 2019). A distinctive feature of the model is that it enables to capture the idea that fairness evaluations are multidimensional: an idea that is well established in the behavioural science literature (see e.g. Kahneman (1992) and Ordóñez et al. (2000)), but it is rarely explored in theoretical and applied work in economics. In the model of this paper, the impact of horizontal fairness considerations, which captures the central hypothesis behind the effect of pay inequality on effort, is only one among other possible influences of perceived fairness and morale: being paid less (more) than a co-worker decreases (increases) morale, effort becomes more (less) costly, and decreasing (increasing) effort will increase workers' utility. The model is also suitable to analytically characterise workers' incentive for reciprocity in the context of gift-exchange games—according to which low (high) wages are reciprocated by low (high) effort—and the incentive for income targeting in the context of piece-rate payment schemes, where the incentive to generate income through effort is stronger—due to loss aversion—if workers believe their income is below a certain (subjective) target.

The assessment of this paper has produced a number of insights. For instance, the analysis shows that unless the information about the employer's wage-choice set is common knowledge in control treatments (in which pay is secret, or subjects are equally paid), experiments based on three-person gift-exchange games cannot clearly disentangle whether effort responses are driven by horizontal fairness concerns, reciprocity towards the employer, or a combination of both. This is in fact the sort of concern which motivated the use of piece-rate payments in laboratory experiments, without subject-employers choosing pay. The analysis also shows why effort outcomes under piece-rate payments should be interpreted as the result of two offsetting behavioural forces: the morale effect of pay inequality—which may lead to lower effort from lower-paid employees—and the effect stemming from the incentive for income targeting—which may lead to higher effort from lower-paid employees. These considerations also enable to explain the evidence on subject's responses in terms of quality *versus* quantity of effort when faced with disadvantageous pay inequality: since the piece rate depends on quantity, subjects have an incentive to increase effort on that dimension, e.g. increasing the output produced, but forgoing quality, e.g. making more mistakes. Further, the paper also discusses how other features of experimental designs, such as the use of repeated interactions—which could foster reputation building in gift-exchange games—and the role of procedural fairness—which could affect how

pay inequality is perceived, may have confounded or impaired the identification of the effect of pay inequality on effort in existing experiments.

In light of these findings, the paper provides a more nuanced overview of the evidence reported in the literature. While some experiments appear to have managed to control for potential confounds, there are also a number of studies from which it is more difficult to draw definitive conclusions on the effect of pay inequality on effort. As such, this paper calls for a more cautious and informed interpretation of the evidence reported in the literature.[5] Despite an initial hope to obtain a more transparent picture of the effect of pay inequality on effort after the assessment, we conclude that evidence on this hypothesis remains still mixed, and that further research is needed, especially in settings where the experimenter can exert control over the issues highlighted in this paper. In support of this assertion, the paper concludes by delineating a set of recommendations on how to improve the design of experiments in the future.

The paper is organised as follows. Section 2 provides an overview of existing laboratory designs and evidence, and highlights the presence of potential confounds. Section 3 constructs a taxonomy of treatments to classify existing designs, and develops a theoretical model to characterise employees' effort choices in the laboratory. Section 4 applies this conceptual framework to critically assess the literature, while Section 5 discusses the existing evidence in light of this assessment and provides a number of recommendations for the development of future laboratory designs. Section 6 offers some concluding remarks.

2 An overview of existing designs and evidence

This section provides an overview of the most distinctive aspects of laboratory designs that investigate the effect of pay inequality on effort. Importantly, the overview is based on those studies in which pay inequality stems from horizontal pay comparisons between workers, and in which the main outcome of interest is workers' effort. Other laboratory experiments with similar designs that are not concerned with the effect of horizontal pay inequality on workers' effort are, obviously, not discussed.[6]

[5] As such, the paper also suggests caution regarding the use of a meta-analysis to evaluate estimates of pay-inequality treatment effects in the experimental literature. Despite the attempt of this paper to provide a more systematic assessment of existing studies, as it will become clearer in the remainder of the paper there is still a considerable amount of heterogeneity in the way in which experiments are designed, with some featuring even more than one potential confound. These aspects may pose a serious challenge to meta-analyses investigations.

[6] For instance, the laboratory experiments of Maximiano, Sloof, and Sonnemans (2007), Hennig-Schmidt et al. (2010), Abeler, Altmann, Kube, and Wibral (2010), Hesse and Rivas (2015), and Gross, Guo, and Charness (2015), all implement multi-lateral gift-exchange games to investigate the determinants of worker's effort in a

The overview is based around the two most widely used settings in the literature, three-person gift-exchange games and piece-rate payment schemes, and it is further organised along three dimensions: overall design of the experiments; reported evidence; and potential issues.

2.1 Three-person gift-exchange games

Three-person gift-exchange games have been adopted in the experiments of Charness and Kuhn (2007), Gächter and Thöni (2010), Gächter et al. (2012), Nosenzo (2013), and Charness, Cobo-Reyes, Lacomba, Lagos, and Perez (2016).

Overall Design. The design of these experiments features a gift-exchange game between one subject-employer and two subject-employees. In its canonical version, the employer moves first, and chooses the wage of each employees from a given set of options—the employer's choice set; the employees move second, they observe the wage, and choose effort, which in turn affects the employer's payoff. The pay inequality treatment features information on relative wages being public knowledge, and it arises (endogenously) as the outcome of the employer's wage choice. A number of studies use effort outcomes from pay equality as the control treatment (e.g. Gächter and Thöni (2010) and Gächter et al. (2012)), while others use an equivalent three-person gift-exchange game in which information on relative wages is kept private (e.g. Nosenzo (2013)). Variations include: the case in which wages are set exogenously by a random mechanism (as in one treatment of Gächter and Thöni (2010)), or by the experimenter (only for one employee, as in one treatment of Nosenzo (2013)); the case in which the choice of the wage can be delegated to one or both employees (Charness et al., 2016); and the case in which the effort choice of one employee is observable by their peer (as in the main experiment of Gächter et al. (2012)). Other variations include whether the gift-exchange game is played once or in multiple rounds, and whether subjects are randomly rematched if the game is repeated.

Reported Evidence. Evidence from experiments adopting this design is mixed. Charness and Kuhn (2007) find that pay inequality has no effect on employees' effort, while Nosenzo (2013)

multi-worker environment. However, in Maximiano et al. (2007) workers are paid the same wage in all treatments, while Hesse and Rivas (2015) investigate how the employers' pay distribution affect worker's effort decision. Hence, there is no horizontal pay inequality treatment in these studies. Abeler et al. (2010) and Gross et al. (2015) implement designs featuring a three-person gift-exchange game where, importantly, employees are ex-ante different. Both studies aim to investigate the effect of *pay equality* on workers' effort when workers are heterogenous in their productivity (see also the discussion on procedural fairness in Section 4). Rather, Hennig-Schmidt et al. (2010) do not explicitly investigate the effect of pay inequality on effort in the laboratory version of their experiment. Finally, Brandts, Ortiz, and Belda (2019) adopt a 'reverse' multilateral gift-exchange game, where workers choose effort before the employer chooses the wage allocation, and study the employer's wage decision as the outcome of interest.

and Gächter et al. (2012) find some evidence of significant negative effects of disadvantageous pay inequality on effort. Also Gächter and Thöni (2010) find evidence that employees' effort is decreasing in their co-worker's wage, but this effect disappears in their treatment in which the subject-employer is passive and wages are set randomly. Finally, Charness et al. (2016) report that, under a public information treatment, both advantageous and disadvantageous pay inequality *increase* workers' effort relative to the case of pay equality[7]

Potential Issues. Three-person gift-exchange games have the advantage of being more representative of real-world employment relationships, in which the internal pay structure is chosen by the employer. However, in this setting it may be harder to identify whether it is the morale effect of pay inequality—stemming from horizontal fairness considerations—or the reciprocal incentive towards the subject-employer—i.e. vertical fairness—that drives the effort choice of employees. For instance, in a situation in which an employee that is paid a relatively low wage exerts low effort, it may not be straightforward to conclude whether the low effort response is triggered by the feeling of being paid less than their peer, or by the employee negatively reciprocating the employer for choosing to pay them the low wage relative to a higher alternative, regardless of how much their peer is paid.

2.2 Piece-rate payment schemes

Piece-rate payment schemes have been adopted in the experiments of Burchett and Willoughby (2004), Bartling and von Siemens (2011), Ku and Salmon (2012), Greiner, Ockenfels, and Werner (2011), Bracha et al. (2015), and Cardella and Roomets (2022).

Overall Design. The design of these experiments generally features subject-employees being paid a piece-rate wage to work on a real-effort task for several rounds or for a given period of time.[8] The inequality treatment is typically administered by dividing subjects into two groups: one receiving a lower piece rate and one receiving a higher piece rate, this distinction being public knowledge. Hence, pay inequality is captured by unequal piece rates rather than unequal flat wages. These rates are typically decided by the experimenter and the assignment of subjects to different piece rates is random. Control treatments are either given by sessions in which both groups receive equal (low or high) piece rates, and this is public knowledge (e.g. Bartling and

[7]Since their design is based on repeated interactions without re-matching, Charness et al. (2016) attribute this finding to the possibility that employees were trying to influence the employers decision in the next round, essentially playing a 'reverse' gift-exchange game. This potential issue, which stems from repeated interactions, will be discussed in Section 4.

[8]The exception is Bartling and von Siemens (2011), in which participants have to state their effort choice.

von Siemens (2011) and Ku and Salmon (2012)), or by sessions in which the information on piece rates differential is kept private (e.g. Burchett and Willoughby (2004) and Bracha et al. (2015)).

Reported Evidence. The evidence reported by these class of experiments also appears to be mixed. Burchett and Willoughby (2004) find that subjects who were paid the low (high) piece rate decrease (increase) their effort when information of relative wages is public knowledge. Also, Ku and Salmon (2012) and Bracha et al. (2015) find some evidence that pay inequality affects effort, but only for the case of disadvantageous inequality. On the other hand, Bartling and von Siemens (2011) cannot reject the null hypothesis that pay inequality has no effect on employees' effort. Finally, the experiments of Greiner et al. (2011) and Cardella and Roomets (2022) report evidence on employees' effort in terms of both quantity and quality.[9] While both studies report that disadvantageous (piece-rate) pay inequality had a negative effect on the quality of effort, Cardella and Roomets (2022) find no effect on the quantity of output produced, while Greiner et al. (2011) find that employees' effort on this dimension significantly increased.

Potential Issues. Designs adopting piece-rate payments enable to avoid the existence reciprocal responses towards an employer (a point firstly made by Bartling and von Siemens (2011)). However, piece-rate payment schemes introduce another potential confound if subjects are "income targeters". Income targeting behaviour reflects the idea that workers have an incentive to exert more effort, and therefore generate higher earnings under piece rates, until they reach a (subjective) income target (Goette et al., 2004; Fehr and Goette, 2007). In the context of the experimental designs discussed above, if the low-rate employee perceives their pay to be below their income target, which could be influenced by the expected final earnings of their high-rate peer, disadvantageous pay inequality may generate an incentive to increase effort, in order to reduce the expected final earnings gap. In such a case, the income targeting incentive could dampen, offset, or even dominate any potential effect of disadvantageous (piece-rate) inequality on effort.

Overall, experimental laboratory evidence in favour of the hypothesis that horizontal pay comparisons matter for employees' perceptions of fairness and effort appears to be inconclusive.

[9] In both experiments the piece-rate payment depends on the quantity of output produced (respectively, number of forms completed / applications submitted) and not on the quality employees' effort (share of correctly transcribed forms / applications). The key implications of this observation will become clearer to the reader in the remainder of the paper.

In summary, while some studies find no evidence in support of this hypothesis, other studies find that only disadvantageous pay inequality has a significantly stronger and negative impact on employees' effort. This overview also highlighted the existence of potential confounding factors, as well as other nuances, that may have impaired the identification of treatment effects in existing laboratory designs, perhaps contributing to the evidence being mixed. The next two sections aim to provide a more formal assessment of this conjecture.

3 Conceptual framework

This section provides a conceptual framework that will constitute the common ground on which existing experimental designs will be assessed. The following section develops a taxonomy of treatments and testing approaches which can be used to classify, and more clearly compare, existing designs along the dimensions discussed above. Then, the subsequent section will develop a model of worker preferences and effort choice which will be adopted to conduct a more formal assessment.

3.1 A taxonomy of treatments and testing approaches

To begin, it is useful to summarise the objective of the literature discussed in this paper within the following hypothesis testing framework:

H0: *Pay inequality has no effect on employees' effort.*

H1: *Pay inequality affects employees' effort: disadvantageous (advantageous) pay inequality has a negative (positive) effect on employees' effort.*

Hence, the literature seeks to identify the effect of pay inequality—stemming form horizontal fairness considerations—on employees' effort.

Next, let us classify the treatments and controls adopted by the existing designs to test this hypothesis into five main dimensions. The first is pay determination, which can be "endogenous", as in gift-exchange games in which it is a participant who chooses wages, or "exogenous", as in designs in which wages are manipulated by the experimenter or set randomly. There are also two types of pay: a "flat wage", and/or a "piece rate". Subjects are also typically divided between two main types depending on the level of their pay: those receiving a relatively "high" wage and those receiving a relatively "low" wage. Sessions can also differ in terms of relative

pay: under "equality" treatments, subjects are paid the same wage; while under "inequality" treatments, subjects in the same match/group/team are paid different wages. Finally, treatments can differ depending on whether information on relative pay is "private" or "public". These treatment dimensions are summarised in the table below.

Table 1: Treatment Dimensions

Pay determination	endogenous; exogenous
Pay type	flat wage; piece rate
Pay level	high; low
Relative pay	equality; inequality
Information	public; private

Finally, it can be useful to summarise more formally the different empirical strategies adopted by the experimental literature. In fact, while some studies used average effort outcomes under a private-inequality treatment as a control for the effect of pay inequality on effort, others used average effort outcomes under a public-equality treatment as the relevant control. Clearly distinguishing between these and other testing approaches is important, as different empirical strategies could, potentially, lead to different interpretations of treatments effects and related evidence. To do so, consider two workers, i and j, and denote their effort by e and the relevant pay type (flat wage or piece rate) by x. It is then possible to summarise the testing approaches adopted by the existing literature as follows:

$$\textbf{H1-A.} \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] \gtrless \mathbb{E}\left[e_i|\text{private; inequality}\right] \quad \Leftrightarrow x_i \gtrless x_j;$$

$$\textbf{H1-B.} \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] \gtrless \mathbb{E}\left[e_i|\text{public; equality}\right] \quad \Leftrightarrow x_i \gtrless x_j;$$

$$\textbf{H1-C.} \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] \gtrless \mathbb{E}\left[e_j|\text{public; inequality}\right] \quad \Leftrightarrow x_i \gtrless x_j;$$

where, for instance, the observation that $\mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{private; inequality}\right]$ when $x_i < x_j$ would be interpreted as evidence in favour of the hypothesis that disadvantageous pay inequality has a negative effect on employees' effort. The only exception are Bartling and von Siemens (2011) who in fact tested for:

$$\textbf{H1-B'.} \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] \neq \mathbb{E}\left[e_i|\text{public; equality}\right] \quad \Leftrightarrow x_i \neq x_j.$$

Next, we can begin to use this taxonomy of treatments and testing approaches to classify existing laboratory experiments. For instance, the experiments of Gächter and Thöni (2010)

and Nosenzo (2013) are both based on a three-person gift exchange game, however, the design of Gächter and Thöni (2010) only features a public information treatment with a testing approach based on the comparison of effort outcomes between the equality and inequality treatment; while the design of Nosenzo (2013) features both public and private information treatments and his testing approach is based on the comparison between these two when employees are facing pay inequality. Hence, the design classification for Gächter and Thöni (2010) is: {endogenous; flat wage} with {public; equality/inequality} treatments; their empirical strategy adopts the {public; equality} as the control treatment, and their hypothesis testing approach is given by H1-B. While for Nosenzo (2013) the classification is: {endogenous; flat wage} with {private/public; inequality} treatments; the control treatment is {private; inequality}, and his testing approach is characterised by H1-A.

Along the same lines we can classify the experiments of Bartling and von Siemens (2011) and Bracha et al. (2015), which are both based on the use of piece-rate payment schemes, but adopt different designs and testing approaches. In fact, while Bartling and von Siemens (2011) implement a public information treatment and compare the effort choices of employees between the pay equality and pay inequality treatments, Bracha et al. (2015) implement both public and private information treatments, but then use average effort from the higher-paid employees under public information as the main control for the response of lower-paid employees. Hence, the design classification for Bartling and von Siemens (2011) is: {exogenous; piece rate} with {public; equality/inequality} treatments; they adopt the {public; equality} treatment as the main control, and their testing approach is better described by H1-B'. While for Bracha et al. (2015) the classification is: {exogenous; piece rate} with {private/public; inequality} treatments; their control is the {public: low; inequality} treatment, and their testing approach is therefore given by H1-C.

This simple classification exercise can be extended to all the other experiments discussed in Section 2. Pursuing this approach will turn out to be particularly useful to enable a more transparent comparison of existing designs and testing frameworks, and to highlight presumably important methodological differences among them. A complete classification of the literature using this taxonomy is presented in Tables 2 and 3 of Section 4.

3.2 A model of worker preferences

This section develops a model of worker preferences that aims to analytically characterise the impact of pay inequality (horizontal fairness), reciprocity (vertical fairness) and income targeting behaviour on a worker's utility and choice of effort. To do so, the modelling approach draws from the literature on distributional preferences (e.g. Fehr and Schmidt (1999) and Bolton and Ockenfels (2000))—according to which it is the distribution of outcomes that matters for fairness considerations; from the literature on intention-based reciprocity (e.g. Rabin (1993) and Dufwenberg and Kirchsteiger (2004))—according to which it is the intention leading to the distribution of outcomes that matters for fairness considerations; and from the literature on reference dependence and loss aversion in the evaluation of pay—see, for instance, Goette et al. (2004) in a context of piece rates, and Dickson and Fongoni (2019) in a worker-firm relationship with flat wages.

Consider a worker i engaged in an employment relationship with an employer and a co-worker j. The job consists in performing a productive task which requires effort e. The employment contract is incomplete: effort is not observable by neither the employer nor their co-worker, meaning that there is some discretion on how much effort workers can exert. Effort is costly to exert, but there are also psychological benefits of being productive, reflecting the existence of intrinsic motivation that is independent of pay. To capture these features, we assume workers perceive some net-cost of effort given by $d(e) = e^2/2 - be$, where $\bar{e} \equiv b \in \mathbb{R}_{++}$—the minimiser of d—captures "normal effort", i.e. effort that is independent of pay.[10]

Both workers choose their effort once they learn the wage ω that they are paid for the task. The wage can either be a flat wage $\omega = w$, or a piece-rate wage $\omega = ez$, where z is the piece rate. Further, it is assumed that workers are reference dependent and loss averse (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991) and that evaluate their pay in relation to a reference point r, according to a piece-wise linear gain-loss function $\mu(\omega - r)$, where $\mu(x) = x$ if $x \geq 0$ and $\mu(x) = \lambda x$ if $x < 0$; in which $\lambda > 1$ captures a worker's degree of loss aversion. Importantly, these evaluations are made on three key domains: one reflecting the effect of relative pay on morale (horizontal fairness); one reflecting the role of fair treatment from their employer (vertical fairness); and one reflecting the effect of relative income on morale under

[10] See also Sliwka and Werner (2017), Macera and te Velde (2018), and Dickson and Fongoni (2019) for other models of worker's preferences featuring a similar assumption. Note this assumption also reflects the evidence that workers engage actively in productive tasks even absent a monetary incentive to do so (Bewley, 2007; Altmann, Falk, Grunewald, and Huffman, 2014).

piece rates. In what follows, we will first describe how each one of these domains affects a worker's utility. Subsequently, we will present how they enter the utility function and comment on their relevance depending on the contract type and working environment.

Denote by r^S the reference point used for horizontal fairness considerations, and assume that it is determined by the co-worker's pay, if this is known. The impact of morale stemming from horizontal fairness considerations (i.e. pay inequality) on worker i's utility is given by:

$$S(e_i, \omega_i, r_i^S) \equiv \begin{cases} e_i \mu(w_i - w_j) & \text{if } \omega = w \\ e_i \mu(z_i - z_j) & \text{if } \omega = ez. \end{cases}$$

This morale component of utility captures the central hypothesis behind the effect of pay inequality on effort: being paid more than a co-worker increases morale, effort becomes less costly, and increasing effort will increase utility; while being paid less than a co-worker decreases morale, effort becomes more costly, implying that utility is increased by reducing effort. Moreover, due to loss aversion, the impact of disadvantageous inequality on morale is larger than the impact of advantageous inequality (see also Fehr and Schmidt (1999)).[11]

Next, denote by r^G the reference wage used for vertical fairness considerations and assume that it is determined by a worker's belief of what the employer should pay them (in the spirit of intention-based reciprocity theory), for a given set of possible pay levels. The impact of morale stemming from vertical fairness considerations (i.e. gift-exchange) on worker i's utility is captured by:

$$G(e_i, \omega_i, r_i^G) \equiv \mu(\omega_i - r_i^G)\phi(e_i, \bar{e}_i)$$

where $\phi(e_i, \bar{e}_i) = e_i - \bar{e}_i$. This reciprocity component of utility is modelled in the spirit of Rabin (1993) and Dufwenberg and Kirchsteiger (2004): ϕ would capture a measure of the worker's kindness towards the employer and μ would capture a measure of the employer's kindness towards the worker: a wage above what the worker believes to be the fair wage is perceived as a gift, and the worker will increase utility by increasing effort (above normal effort $\bar{e}$), which in turn increases the employer's profit (a kind action towards the employer); a wage below what the worker believes the employer could pay them is perceived as unkind, hence the worker will increase utility by decreasing effort (below normal effort $\bar{e}$), reducing the employer's profit

[11] However, note that Fehr and Schmidt (1999) model fairness as "inequity aversion": people dislike inequitable outcomes, independently of whether they are advantageous or disadvantageous for them. This is in contrast with how we model the implications of horizontal fairness considerations, according to which advantageous (disadvantageous) inequality has a positive (negative) effect on morale, and therefore on effort.

(an unkind action towards the employer).[12] Loss aversion implies that negative reciprocity is stronger than positive reciprocity (see also Dickson and Fongoni (2019)).

Finally, we also want to model the possibility that workers are "income targeters" when working under piece rates. We denote the relevant reference income by r^I, which is assumed to be determined by the expected income that will be earned by the co-worker, if information on relative piece rates is public. This assumption reflects the existence of horizontal social comparison on the dimension of income (final earnings), in addition to the one on the dimension of pay. The impact of income targeting behaviour on worker i's utility is given by:

$$I(e_i, \omega_i, r_i^I) \equiv \mu(e_i z_i - r_i^I).$$

This component of utility is modelled in the spirit of Goette et al. (2004), and captures the idea that workers compare their income relative to a subjective reference target, and that the incentive to generate income through effort is stronger—due to loss aversion—if workers believe their income is below this target. In the context of the present framework this incentive implies that, if a worker adopts the expected final earnings of their co-worker as a reference, disadvantageous pay inequality may generate an incentive to increase effort, in order to reduce the expected inequality gap.

A worker i's utility is assumed to be additively separable along these three fairness dimensions just described:

$$u(e_i, \omega_i, r_i^S, r_i^G, r_i^I) = [1-\gamma_i]\omega_i - d(e_i) + \alpha_i S(e_i, \omega_i, r_i^S)\mathbf{1}^S + \beta_i G(e_i, \omega_i, r_i^G)\mathbf{1}^G + \gamma_i I(e_i, \omega_i, r_i^I)\mathbf{1}^I; \tag{1}$$

where the parameters $(\alpha_i, \beta_i, \gamma_i) \in \mathbb{R}^3_+$ capture the subjective weights of these three dimensions on a worker's utility,[13] and $(\mathbf{1}^S, \mathbf{1}^G, \mathbf{1}^I)$ are indicator functions with the following properties: $\mathbf{1}^S = 1$ if information on relative pay is public, and $\mathbf{1}^S = 0$ otherwise (also note that in this latter case r_i^I remains undetermined); $\mathbf{1}^G = 1$ if the employer choice set is known, and $\mathbf{1}^G = 0$ otherwise (note this will also be the case in the absence of an employer choosing pay);[14] and

[12] In Appendix A.1 we show how $G(e_i, \omega_i, r_i^G)$ is analogous to the "reciprocity payoff" stemming from a canonical gift-exchange game with reciprocity á la Dufwenberg and Kirchsteiger (2004), which explicitly accounts for the role of beliefs in the formation of the reference point for fairness evaluations.

[13] Note that the income targeting dimension has been added to the worker's utility in a way that is complementary to the typical incentive given by the piece-rate wage.

[14] Note that if the employer's choice set is unknown, workers will not have enough information to assess the intentionality of the employer's choice, and neither they can know whether the employer could have paid them a different wage than the one they receive. Information about the employer's choice set, and hence about the

$\mathbf{1}^I = 1$ under piece rates, i.e. if $\omega = ez$, and $\mathbf{1}^I = 0$ if pay is determined by a flat wage, i.e. if $\omega = w$.[15] Hence, after learning their pay $\omega = \{w, ez\}$ and for given $r = \{r^S, r^I, r^G\}$, depending on the type of employment contract and working environment, workers will choose the level of effort e that maximises their utility, as given by (1).

This formulation of utility enables us to analytically characterise the impact of horizontal pay comparison, reciprocity, and income targeting, on a worker's perception of fairness of their pay, and therefore on their choice of effort. By modelling fairness as being reference dependent within a framework of multiple reference points, the model of this paper embraces the idea that fairness evaluations can be multidimensional. In fact, recognising that reference points are typically artifacts of economics models, this modelling approach is consistent with the idea that there exist a variety of factors affecting workers' perception of fairness of their pay, and it enables us to capture under a unique framework their relative impact on utility and effort choices.[16]

4 Assessment

This section implements the conceptual framework just developed to formally assess whether the potential issues highlighted in the overview of Section 2 are a concern for the identification of the effect of pay inequality on worker effort in existing laboratory designs.

The assessment will be restricted to the disadvantageous inequality case only. This choice is justified by two reasons. First, this is the only case, with the exception of Burchett and Willoughby (2004) and Charness et al. (2016), for which some evidence of treatment effects was detected in the existing literature, and it is therefore important to assess the validity of such identification in detail. Second, the insights stemming from the analysis of disadvantageous inequality will be fairly similar to what can be concluded from an analogous analysis of advantageous inequality. Hence, to ease the expositional burden, the assessment will henceforth focus

employer's payoff, is a fundamental premise of intention-based reciprocity theories (see, in particular, the findings in Hennig-Schmidt et al. (2010)). In the absence of this information, it is not possible to make vertical fairness evaluations, which justifies why $\mathbf{1}^G = 0$ in this case.

[15] This setting implies that there is no incentive for income targeting if the worker is paid a flat wage. This is obvious, since under flat wages: i) a worker's income is either the wage itself, or the sum of the flat wages received in each round, if there are multiple employment periods; and ii) the worker has no control over the flat wage, which is chosen by the employer. Hence there is no action that the worker can take to affect their final earnings and achieve a desired income target.

[16] In particular, the modelling approach taken in this paper is akin to reference-dependent models in which individuals evaluate outcomes relative to multiple reference points according to a 'segregated' mechanism (Kahneman, 1992; Ordóñez et al., 2000)

on disadvantageous pay inequality.

4.1 Gift exchange and reciprocity

First, we classify in Table 2 below the gift-exchange games previously discussed using the taxonomy developed in Section 3. We can then assert that the literature has adopted two testing

Table 2: Three-person Gift-Exchange Games

	Experimental Design	*Control Treatment*	*Testing Approach*	*Evidence (Reported)*
Charness and Kuhn (2007)	endogenous flat wage private/public equality/inequality	{private; inequality} {public; equality}	H1-A H1-B	No
Gächter and Thoni (2010)	endogenous /exogenous flat wage public equality/inequality	{public; equality}	H1-B	Yes (disadvantageous)
Gächter et al. (2012)	endogenous flat wage public equality/inequality	{public; equality}	H1-B	Yes (disadvantageous)
Nosenzo (2013)	endogenous flat wage private/public inequality	{private; inequality}	H1-A	Yes (disadvantageous)
Charness et al. (2016)	endogenous flat wage private/public equality/inequality	{public; equality}	H1-B	Yes* (advantageous)

*Note: Charness et al. (2016) find that both advantageous and disadvantageous pay inequality have a positive effect on effort.

approaches. One comparing average effort outcomes between a private information treatment and a public information treatment, both featuring pay inequality (H1-A); and one comparing average effort outcomes between a pay equality treatment and a pay inequality treatment, both featuring public information on relative wages (H1-B). In the context of the model developed in this paper, failure to reject the null hypothesis H0 would suggest that $\alpha_i = 0$; while rejection of the null hypothesis in favour of the alternative H1 would suggest that $\alpha_i > 0$. However, it is not clear that by investigating H1-A or H1-B, these approaches are able to clearly disentangle the effect of horizontal comparison, on employee's effort, from the one triggered by a reciprocal response towards the employer.

To see this, consider the optimal effort choice of a worker i for any given ω_i, ω_j and r_i^G in a

three-person gift-exchange game. This is given by the value of effort that maximises (1) when $\omega = w$ and $\mathbf{1}^I = 0$:

$$e_i = \bar{e}_i + \alpha_i \mu(w_i - w_j)\mathbf{1}^S + \beta_i \mu(w_i - r_i^G)\mathbf{1}^G. \tag{2}$$

Next, denote the low and high wage by w^L and w^H respectively, where $w^L < w^H$, and assume that, in treatments in which the employer's choice set is common knowledge, the reference wage used for vertical fairness considerations is given by the highest possible wage that the employer can choose: $r_i^G = w^H$. Although being a simplification, this assumption remains in the spirit of existing theories of intention-based reciprocity, since it implies that r_i^G depends on the worker's beliefs of what wage the employer could have chosen (in this case, the highest available alternative).[17] Finally, we assume $w_i = w^L$ and $w_j = w^H$ in the inequality treatment, and $w_i = w_j = w^L$ in the equality treatment.

Let us begin by considering the case in which the employer's choice set is unknown in control treatments (i.e. $\mathbf{1}^G = 0$ in the {private; inequality} and the {public; equality} treatments). Using the optimal effort function (2), our model yields the following average effort outcomes in each treatment (henceforth, coefficients without subscripts indicate average values):

$$\mathbb{E}\left[e_i|\text{private; inequality}\right] = \bar{e}, \tag{3}$$

$$\mathbb{E}\left[e_i|\text{public; equality}\right] = \bar{e}, \tag{4}$$

$$\mathbb{E}\left[e_i|\text{public; inequality}\right] = \bar{e} + \lambda[\alpha + \beta][w_i - w_j]. \tag{5}$$

Together, these imply that, when $w_i < w_j$:

H1-A. $\rightarrow$ $\mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{private; inequality}\right] \Leftrightarrow \forall(\alpha,\beta) \in \mathbb{R}^2_+ \neg(\alpha = \beta = 0)$;

H1-B. $\rightarrow$ $\mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{public; equality}\right] \Leftrightarrow \forall(\alpha,\beta) \in \mathbb{R}^2_+ \neg(\alpha = \beta = 0)$.

[17] In the intention-based reciprocity literature r_i^G would be defined as the worker i's belief of the average of the maximum and minimum material payoffs that the employer could give to i in principle, given the employer's belief of what the worker i will do (see Rabin (1993) and Dufwenberg and Kirchsteiger (2004)). Nevertheless, as also Dufwenberg and Kirchsteiger (2004) point out, there is no deep justification for picking the average of the maximum and minimum payoffs, and since we are not considering the employer's optimal choice, it is not necessary to model higher order beliefs (see also Dufwenberg and Kirchsteiger (2000)). Hence, the assumption adopted here becomes instrumental in capturing the central idea behind intention-based reciprocity: if worker i is paid the low wage they would perceive this as an unkind action from the employer, who could have paid them a higher alternative; while if worker i is paid the high wage, they would perceive this action just as fair. See Appendix A.1 for a detailed analysis of how the model of this paper is related to, and can be derived from, a more general sequential-move game with reciprocity.

We can draw two insights. First, there is no difference between adopting {private; inequality} or {public; equality} as the relevant control treatment, since in both cases workers only observe one wage in both treatments, in this case w^L, and do not have enough information to assess whether the employer could have paid them another wage, in this case w^H. Second, it is not possible to conclude that the observation of low effort from low-paid workers in the {public; inequality} treatment is exclusively triggered by horizontal fairness concerns ($\alpha > 0; \beta = 0$); reciprocity towards the employer alone ($\alpha = 0; \beta > 0$); or a combination of both ($\alpha > 0; \beta > 0$). This suggests that experimental designs based on three-person gift-exchange games, in which the employer's choice set is unknown in control treatments, cannot clearly disentangle whether the effort response of subject-employees to pay inequality is the outcome of horizontal social comparison with their matched peers, or simply reciprocity towards the choice of the subject-employer.[18]

In fact, in order to investigate this issue, Gächter and Thöni (2010) conducted an additional experimental session in which the choice of wages is left to a random device which acts on behalf of the subject-employer (and this is public knowledge). What they found is that, contrary to what they observed in the canonical version of the game, wage inequality had no effect on employees' effort. In light of the framework developed in this paper, the findings of Gächter and Thöni (2010) seem to suggest that the negative effect of disadvantageous inequality on effort observed by the authors was mainly driven by reciprocity towards a subject-employer rather than by concerns for horizontal fairness, i.e. it seems to be the case that $\alpha = 0$ and $\beta > 0$.

Next, let us consider an alternative design in which the employer's choice set is common knowledge in all treatments (i.e. $\mathbf{1}^G = 1$). Designs adopting the strategy method, as opposed to the direct response method, could fall into this category (i.e. Gächter et al. (2012) and Nosenzo (2013)). In this case, average effort predicted by our model in both of the control treatments is:

$$\begin{aligned} \mathbb{E}\left[e_i|\text{private; inequality}\right] &= \bar{e} + \beta\lambda[w_i - w^H], \\ \mathbb{E}\left[e_i|\text{public; equality}\right] &= \bar{e} + \beta\lambda[w_i - w^H], \end{aligned}$$

[18] Note that the literature also adopted a regression approach to estimate the impact of the wage difference $[w_i - w_j]$ (e.g. Charness and Kuhn (2007)), or of the co-worker's wage w_j (e.g. Gächter and Thöni (2010)), on worker i's effort. As it is now clear from the equation describing a worker's effort response to wage inequality in (5), these approaches would be estimating the combined effect of horizontal comparison *and* reciprocity on effort, i.e. the coefficient $[\alpha + \beta]$, and not α alone, which instead is what these studies are seeking to measure.

which implies that, when $w_i < w_j$,

$$\textbf{H1-A.} \rightarrow \mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{private; inequality}\right] \Leftrightarrow \alpha > 0;$$

$$\textbf{H1-B.} \rightarrow \mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{public; equality}\right] \Leftrightarrow \alpha > 0.$$

This result suggests that making the employer's choice set common knowledge when information on relative pay is kept private, or when both workers are paid the same wage, can let the experimenter control for the existence of reciprocity behaviour even in control treatments. Using this approach will then allow the experimenter to conclude that any difference in effort outcomes between the private and public information treatments, or between the equality and inequality treatments, is exclusively driven by the morale effect of horizontal fairness considerations (i.e. that $\alpha > 0$). Hence, this approach enables a cleaner identification of the main treatment effect of interest.

4.2 Piece rates and income targeting

Let us now classify the experimental designs adopting a piece-rate payment scheme in Table 3 below. There are three main testing approaches. Two are the same as the ones adopted in the rest of the literature previously discussed (H1-A and H1-B), while the third approach compares average effort outcomes between the higher-paid and the lower-paid employees within the same treatment of public information and wage inequality (H1-C). The only exception is Bartling and von Siemens (2011) who tested for H1-B'. This case will be discussed separately as a postscript to this section. In what follows, the aim is to assess whether these testing approaches can cleanly identify the effect of pay inequality on effort, that is, whether failure to reject the null hypothesis H0 is due to $\alpha_i = 0$, or whether the rejection of the null hypothesis in favour of the alternative H1 is due to $\alpha_i > 0$.

Let us begin by considering the optimal effort choice of a worker i under a piece-rate payment scheme, for any given ω_i, ω_j and reference income r_i^I. This is given by the value of effort that maximises (1) when $\omega = ez$ and when there is no employer (i.e. $\mathbf{1}^G = 0$):

$$e_i = \bar{e}_i + \alpha_i \mu (z_i - z_j)\mathbf{1}^S + [1 - \gamma_i + \mu_i' \gamma_i] z_i, \tag{6}$$

where $\mu_i' = 1$ when the worker perceives their wage to be above their reference income ($\omega_i > r_i^I$);

Table 3: Piece-rates Payment Schemes

	Experimental Design	*Control Treatment*	*Testing Approach*	*Evidence (Reported)*
Burchett and Willoughby (2004)	exogenous piece rate private/public inequality	{private; inequality}	H1-A	Yes
Bartling and von Siemens (2010)	exogenous piece rate public equality/inequality	{public; equality}	H1-B'	No
Ku and Salmon (2012)	exogenous piece rate public equality/inequality	{public; equality}	H1-B	Yes (disadvantageous)
Greiner et al. (2011)	exogenous piece rate private/public equality/inequality	{private; inequality} {public; equality} {public; inequality}	H1-A H1-B H1-C	Yes* (disadvantageous)
Bracha et al. (2015)	exogenous piece rate private/public inequality	{public; inequality}	H1-C	Yes (disadvantageous)
Cardella and Roomets (2022)	exogenous piece rate public equality/inequality	{public; equality}	H1-B	No*

*Note: Greiner et al. (2011) (Cardella and Roomets (2022)) find that pay inequality has a positive (null) effect on effort—measured in terms of quantity of output produced, while both studies find that it negatively affects the quality of the output. In line with how effort has been measured in the rest of the literature (i.e. in terms of quantity of output produced), we report this as some (no) evidence.

and $\mu_i' = \lambda_i$ when the worker perceives their wage to be below it ($\omega_i < r_i^I$). If the information on relative piece rates is public, we follow the premises of Section 3 and assume that a worker i's reference income is given by their expectations of what their co-worker j could earn, that is $r_i^I = \mathbb{E}\omega_j$. Moreover, to proceed with the analysis we further assume that ω_j will be expected to be higher (lower) than their wage, ω_i, if worker j is paid a higher (lower) piece rate, i.e. $\mathbb{E}\omega_j \gtreqless \omega_i$ if $z_j \gtreqless z_i$.[19] On the other hand, if the information on relative piece rates is private, we will adopt an agnostic approach to the way in which a reference income r_i^I is formed, and we will derive results for all possible cases: $\omega_i \gtreqless r_i^I$.[20]

Next, denote the low and high piece rate by z^L and z^H respectively, where $z^L < z^H$, and let us restrict once again the analysis to the disadvantageous inequality case only. Hence, we assume $z_i = z^L$ and $z_j = z^H$ in the inequality treatment, and $z_i = z_j = z^L$ in the equality treatment. Using the effort function in (6), the model yields the following average effort outcomes in each treatment:

$$\mathbb{E}\left[e_i|\text{private; inequality; }\right] = \bar{e} + [1 - \gamma + \mu'\gamma]z_i, \tag{7}$$

$$\mathbb{E}\left[e_i|\text{public; equality}\right] = \bar{e} + [1 - \gamma]z_i, \tag{8}$$

$$\mathbb{E}\left[e_i|\text{public; inequality}\right] = \bar{e} + \alpha\lambda[z_i - z_j] + [1 - \gamma + \lambda\gamma]z_i, \tag{9}$$

$$\mathbb{E}\left[e_j|\text{public; inequality}\right] = \bar{e} + \alpha[z_j - z_i] + z_j. \tag{10}$$

Notice that, unlike the case of gift-exchange games, unless subjects perceive they can achieve their reference target when information is private, the treatments {private; inequality} and {public; equality} may lead to different effort outcomes, which implies these treatments do not always represent the same control.[21]

We can now use the effort functions just derived to analyse the testing approaches adopted

[19] Since $\mathbb{E}\omega_j = \mathbb{E}[e_j z_j]$ where z_j is known, whether this assumption holds crucially depends on a worker i's expectations about how much effort their co-worker j will exert, which in turn will depend on worker j expectations about how much effort i exerts. Since effort choices are made simultaneously, workers will have to form expectations about their co-worker's effort and expectations, and these will have to be consistent with each other and with workers' expected behaviour. In Appendix A.2 we provide a more formal analysis of this situation, and we show that the assumption we make is reasonable if the pay gap $|z_i - z_j|$ is sufficiently large, for given α, γ, and λ.

[20] It is not inconceivable that, after learning that pay is determined through piece rates, subjects begin to form beliefs about a personal income target to be achieved throughout the experiment.

[21] Further notice that these theoretical results and the following analysis are based on the implicit assumption that a worker will always manage to exert their optimal, utility-maximising level of effort within a given experimental setting. However, there may be constraints that limit this possibility, such as a pre-set time to engage with the task. In such cases, it may be harder to detect the variation in effort outcomes triggered by the piece-rate/income-targeting incentive, which further contributes to the uncertainty around the identification of treatment effects under piece rates.

in the literature when $z_i < z_j$. Hence, we find that:

$$\textbf{H1-A.} \rightarrow \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{private; inequality}\right] \quad \Leftrightarrow \quad \gamma < \alpha \frac{\lambda}{\lambda - \mu'} \frac{z_j - z_i}{z_i};$$

$$\textbf{H1-B.} \rightarrow \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_i|\text{public; equality}\right] \quad \Leftrightarrow \quad \gamma < \alpha \frac{z_j - z_i}{z_i};$$

$$\textbf{H1-C.} \rightarrow \quad \mathbb{E}\left[e_i|\text{public; inequality}\right] < \mathbb{E}\left[e_j|\text{public; inequality}\right] \quad \Leftrightarrow \quad \gamma < \frac{\alpha[\lambda + 1] + 1}{\lambda - 1} \frac{z_j - z_i}{z_i}.$$

A careful inspection of these conditions enables us to draw a number of insights. First, our model suggests that disadvantageous piece-rate inequality has a negative effect on workers' effort only if the incentive for income targeting is sufficiently small *relative* to the morale effect of pay inequality. This result implies that the incentive to work harder, in order to reduce the expected disadvantageous income gap, may dampen, offset, or even exceed the negative effect of pay inequality on morale and effort, confounding the identification of the main treatment effect of interest. Hence, laboratory experiments based on piece-rate payment schemes seem to be more appropriately designed to test for the relative importance of the morale effect of pay inequality *versus* income inequality, on effort. This consideration reveals an issue: the absence of differences in average effort outcomes between treatment and control groups cannot be interpreted as evidence in support of the null hypothesis that pay inequality has no effect on effort, since this could also be the outcome of two offsetting forces.[22]

Next, note that if the pay gap $z_j - z_i$ is large enough, it will be more likely that *ceteris paribus* the conditions on the right-hand side of the material equivalence above will hold: by purposely setting a large gap between the low and the high piece rate, the experimenter can improve the ability of these designs to isolate the effect of horizontal fairness considerations from the income targeting incentive. In fact, if the difference in piece rates is large enough, lower-paid employees may feel discouraged to even attempt to increase their effort as a mean to reduce the expected pay gap.

Another possible solution to this issue can be deduced in the context of hypothesis H1-A. In particular, notice that if $\omega_i < r_i^I$ in the {private; inequality} treatment, the condition on the right-hand side of the material equivalence for hypothesis H1-A will be given by $\alpha > 0$.[23] That

[22]The only experiment acknowledging this is the one of Ku and Salmon (2012), who in fact frame their hypotheses accordingly.

[23]This can be deduced from comparing equation (7), in which $\mu' = \lambda$ since $\omega_i < r_i^I$, with equation (9) as required by the testing approach H1-A. From this, it is straightforward to check that $\mathbb{E}\left[e_i|\text{public; inequality}\right] <$

is, if the experiment is designed such that subjects will have an incentive for income targeting in the domain of losses even when information on relative pay is private, then, any observed difference in effort outcomes between the {private; inequality} and the {public; inequality} treatment can be attributed exclusively to the negative effect of pay inequality that stems from horizontal fairness considerations.

The conditions derived above also highlight a number of caveats regarding the testing approaches adopted in the literature. For instance, when testing hypothesis H1-A, it should be acknowledged that the more a worker is loss averse, the stronger would be the incentive to increase effort in order to reduce the disadvantageous pay gap—perceived as a loss. Hence, the more a worker is loss averse, the more likely it will be that the positive effect of income targeting on effort offsets the negative effect of piece-rate inequality on morale and effort, again confounding the identification of the latter effect. As such, gathering information on subjects' degree of loss aversion can be useful to control for this when testing for hypothesis H1-A under piece rates. Finally, by inspecting the condition on the right-hand side of the material equivalence concerning hypothesis H1-C, it can be deduced that this condition holds even if $\alpha = 0$, that is, even if employees do not care about horizontal fairness and inequality in piece rates. This deduction implies that the average effort of higher-paid employees under public information is not an appropriate control to test for the effects of pay inequality on effort. Under piece rates effort is an increasing function of pay: the piece rate effectively captures the marginal monetary benefit of exerting an additional unit of effort. Hence, it is possible that higher-paid workers will exert more effort than lower-paid workers, regardless of horizontal fairness considerations and inequality.

Finally, let us consider hypothesis H1-B'. Using the effort function in (6) when $z_i \neq z_j$ yields:

$$\mathbb{E}\left[e_i|\text{public; inequality}\right] = \bar{e} + \alpha\mu'[z_i - z_j] + [1 - \gamma + \mu'\gamma]z_i;$$

which implies that:

$$\textbf{H-B'.} \rightarrow \mathbb{E}\left[e_i|\text{public; inequality}\right] \neq \mathbb{E}\left[e_i|\text{public; equality}\right] \quad \Leftrightarrow \quad \gamma \neq \alpha\frac{z_j - z_i}{z_i}.$$

In their study Bartling and von Siemens (2011) conclude that wage inequality has no effect on employees' effort, since they find no significant difference between effort outcomes across

$\mathbb{E}\left[e_i|\text{private; inequality}\right]$ only if $\alpha > 0$.

their treatment and control group. The above deduction, instead, suggests that an alternative, and equally plausible explanation for this evidence would be that the effect of pay inequality on morale and effort is offset by the incentive for income targeting. It is therefore difficult to conclude anything about the effect of pay inequality on employees' effort from this test.

In conclusion, the analysis of this section highlighted the existence of a number of potential confounding factors in designs based on piece rates, and calls for a more cautious interpretation of the evidence. In general, experiments using piece-rate payments seem to lend themselves more naturally to testing for the morale effect of pay inequality *relative* to the effect of income inequality.

4.3 Further considerations

In this section we conclude the assessment of the experimental literature by discussing more loosely about other elements of existing designs that should be considered.

Repeated Interactions and Reputation Building. In the context of three-person gift-exchange games it could be important to consider the impact of repeated interactions on the behaviour of employees. In this respect, experiments vary considerably between designs in which the game is played only once (Nosenzo, 2013), or in multiple rounds; and whether the employees-employer match remains the same in each round (Charness et al., 2016), or whether subjects are randomly re-matched (Charness and Kuhn, 2007; Gächter and Thöni, 2010). Repeated interactions with the same employer may introduce incentives among employees to build a reputation for being "highly productive": employees might increase effort when receiving a low wage, in order to persuade the employer to pay them a higher wage in subsequent rounds (see also the discussion in Charness et al. (2016)). The same incentive could exist even when subjects are randomly rematched in each round (for instance, employees may exert higher effort in one round with the expectation that, if re-matched with the same employer, they will be paid a higher wage in return). Hence, the presence of this incentive may swap the timing of the canonical gift-exchange game, effectively becoming a 'reverse' gift-exchange game with employees as first movers, confounding the identification of the effect of pay inequality on effort.

Justified Inequality and Procedural Fairness. The key premise which motivates the research discussed in this paper is that pay inequality is perceived as unfair by employees, especially by those that feel underpaid relative to their peers. However, the literature on procedural fairness points out that whether inequality is perceived as unfair may depend on the procedure

that generated it, rather than the inequality itself (Bolton, Brandts, and Ockenfels, 2005). This consideration bears important implications for the design of experiments investigating the effects of pay inequality. For instance, consider the experiment of Charness and Kuhn (2007), in which employees are *ex-ante* different in terms of their productivity, this difference being common knowledge. This situation would justify the decision of employers to pay the low wage to the least productive employees and the high wage to the most productive ones. When receiving the low wage, lower-paid employees may infer they are the least productive and perceive the disadvantageous inequality as fair, with no (or less) adverse consequences on morale and effort. As other authors have already pointed out (e.g. Gächter and Thöni (2010), Bartling and von Siemens (2011) and Nosenzo (2013)) this is perhaps why Charness and Kuhn (2007) do not find any evidence of the effect of pay inequality on employees' effort.[24] Thus, if experiments are designed under the premise that unequal allocations are naturally perceived as unfair, it is important to ensure that employees perceive the procedure that generates pay inequality as unjustified, or else, to ensure that this matter is appropriately controlled for.

There are three laboratory experiments in this literature that have tackled the theme of procedural fairness explicitly. One is the design of Ku and Salmon (2012), in which subject-employees are ex-ante equal, and wages are set by a random mechanism. They find that those who viewed the random mechanism as unfair responded more strongly to disadvantageous pay inequality. The other two are the experiments of Bracha et al. (2015) and Cardella and Roomets (2022), in which subjects perform an assessment task—before the start of the main experiment—and are subsequently divided in two sub-treatments: one in which subjects' pay in the main experiment depends on their performance in the assessment task—i.e. the allocation of pay is justified by a *meritocratic* criterion; and one in which subjects' pay is determined by an arbitrary/random criterion—i.e. the allocation of pay is *unjustified*. As such, in both these experiments ex-ante differences between subjects are generated endogenously. Bracha et al. (2015) find that the effect of pay inequality on labour supply is dampened when the allocation of pay is perceived as justified by subjects.[25] In the same vein, Cardella and Roomets (2022) find stronger effects of pay inequality on effort under the unjustified pay inequality treatment. These findings suggest that procedural fairness is an important aspect to consider and neglecting

[24] Other laboratory experiments studying a design in which employees are ex-ante different are those of Güth, Königstein, Kovács, and Zala-Mezo (2001), Abeler et al. (2010), Angelova, Güth, and Kocher (2012) and Gross et al. (2015). See Nosenzo (2013) for a more thorough discussion of this literature.

[25] However, the authors do not perform an analogous analysis on their measure of employees' effort.

its role can be misleading when interpreting the evidence.[26]

The Role of Loss Aversion. In the theory developed in this paper loss aversion is the psychological source of two important behaviours. First, it determines the strength of a worker's effort response to disadvantageous pay inequality. To this extent, loss aversion might explain why the existing literature identified significant treatment effects only for the case of disadvantageous inequality. To see this, denote by $T^{-(+)}$ the disadvantageous (advantageous) inequality treatment and by C the respective appropriate control treatment (and assume there are no confounds). Further denote by $\bar{\tau}^{-(+)}$ the absolute value of the average treatment effect under disadvantageous (advantageous) pay inequality, where $\bar{\tau}^{-(+)} \equiv \left|\mathbb{E}[e_i|T^{-(+)}] - \mathbb{E}[e_i|C]\right|$, and by δ the absolute value of the minimum average treatment effect that an experiment is meant to be able to detect. Using the model of Section 3 to theoretically derive average effort outcomes in each treatment yields: $\mathbb{E}[e_i|T^+] = \bar{e} + \alpha[w^H - w^L]$; $\mathbb{E}[e_i|T^-] = \bar{e} + \alpha\lambda[w^L - w^H]$; and $\mathbb{E}[e_i|C] = \bar{e}$; which implies that:

$$\text{a)} \quad \bar{\tau}^- > \bar{\tau}^+ \Leftrightarrow \lambda > 1 \quad \text{and}$$

$$\text{b)} \quad \exists \lambda > 1 \quad \text{such that} \quad \bar{\tau}^- > \delta > \bar{\tau}^+ \quad \forall \alpha > 0.$$

That is, loss aversion implies that a) the effect of disadvantageous inequality on effort is stronger, and therefore it might explain why b) this effect is more easily identifiable in existing experiments, despite horizontal social comparison in the dimension of pay, that is α, being equally important. Second, loss aversion determines the strength of the incentive to exert more effort, by under-paid subjects, in a situation of expected earnings inequality under piece rates. Both considerations imply that to gain a better understanding of the effects of pay inequality on effort it might be relevant to measure and control for subjects' degree of loss aversion, and to reconsider hypothesis testing approaches that rely on the assumption of symmetric treatment effects between disadvantageous and advantageous inequality.

Quantity versus Quality Tradeoff. In existing experiments adopting piece rates and real-effort tasks, the piece-rate payment depends on the quantity of output produced, but not on its quality. Since quality is not monitored and it does not affect the final payoff from the experiment, it is natural for subjects to put effort to increase the quantity of output, rather than to increase

[26] An extension of the logic of procedural fairness can lead to the assertion that even pay equality could be perceived as unfair, if employees are in fact unequal (see Abeler et al. (2010) and Gross et al. (2015) for laboratory experiments investigating this hypothesis).

its "quality". For example, in the experiment of Greiner et al. (2011), subjects could increase their final earnings by transcribing a greater number of forms (copying 10-digit numbers into an input mask), regardless of whether this was done correctly or not—i.e. regardless of quality. In the analysis of piece-rate payment schemes, we formally established that the incentive to increase effort along the quantity dimension might be particularly salient for the lower-paid subjects, who are even more motivated (loss aversion) to increase their final earnings to reduce the expected inequality gap in income. This theoretical result enables to put some clarity around the puzzling evidence reported by Greiner et al. (2011) (and Cardella and Roomets (2022)) who find that disadvantageous piece-rate inequality has a positive (null) effect on effort, measured in terms of quantity of output, but a significantly negative effect on quality, measured in terms of incorrectly transcribed forms (submitted applications). These considerations suggest that experimental designs should explicitly consider both these dimensions of output—quantity and quality—when reporting evidence on the effect of pay inequality on effort, and the fact that, under piece rates, subjects have an incentive to allocate their effort to increase quantity at the expense of quality.

5 Discussion

The assessment of the previous section has shown that behavioural factors such as reciprocity towards the employer, income targeting, reputation building, and procedural fairness can either strengthen or dampen the effect of pay inequality on employees' morale and effort. If not appropriately controlled for, these issues may impair the identification of treatment effects. Nevertheless, the assessment has also highlighted how a number of existing laboratory designs have managed to avoid these potential confounds. In this section, we will use the key findings from this assessment to shed some light on the evidence reported in the literature. Next, we will draw a number of suggestions on how to improve the design of laboratory experiments that aim to test for the effect of pay inequality on effort in the future.

5.1 Does pay inequality affect worker effort?

In light of the assessment of this paper, it can be useful to re-interpret the evidence on the pay inequality hypothesis that is reported in the literature. The key findings from the assessment are summarised in Table 4, which lists, for each experiment discussed, how the evidence has been

reported in the literature, how the evidence should be reinterpreted in light of the assessment of this paper, what is the main potential issue affecting the identification of treatment effects, and whether the experimental design managed to control for it.

Table 4: Evidence of the Effect of Pay Inequality on Effort

	Evidence		*Issues*	
	Reported	Post-assessment	Potential	Resolved
Gift-exchange Games				
Charness and Kuhn (2007)	No evidence	Inconclusive	Procedural fairness	✗
Gächter and Thoni (2010)	Disadvantageous	No evidence	Reciprocity	✗
Gächter et al. (2012)	Disadvantageous	Disadvantageous	Reciprocity	✓
Nosenzo (2013)	Disadvantageous	Disadvantageous	Reciprocity	✓
Charness et al. (2016)	Advantageous	Inconclusive	Reputation building	✗
Piece-rate Payment Schemes				
Burchett and Willoughby (2004)	Some evidence	Relative evidence	Income targeting	✗
Bartling and von Siemens (2010)	No evidence	Inconclusive	Income targeting	✗
Ku and Salmon (2012)	Disadvantageous	Relative evidence	Income targeting	✓
Greiner et al. (2011)	Some evidence*	Relative evidence	Income targeting	✗
Bracha et al. (2015)	Disadvantageous	Inconclusive	Testing approach	✗
Cardella and Roomets (2022)	No evidence*	Relative evidence	Income targeting	✗

*Note: Recall that Greiner et al. (2011) (Cardella and Roomets (2022)) find that pay inequality has a positive (null) effect on effort—measured in terms of quantity of output produced, while both studies find that it negatively affects the quality of the output. In line with how effort has been measured in the rest of the literature, we interpret this as some (no) evidence pre assessment, and relative evidence post assessment.

Unfortunately, the picture stemming from Table 4 is not much clearer than the one at the start of our assessment. Nevertheless, we can draw some general observations. In terms of experiments adopting three-person gift-exchange games, the evidence remains mixed, even though a number of designs managed to control for the potential issue of vertical fairness towards a subject-employer—either by removing intentionality (Gächter and Thöni, 2010), or by allowing employees to know the employer's choice set (Gächter et al., 2012; Nosenzo, 2013). On the other hand, the interpretation of the evidence from experiments using piece rates requires some caution. First, as a general principle, whatever is the effect of pay inequality on effort that is documented in these sttudies, it should be considered as evidence relative to the effect of income targeting on effort. Second, it is important to acknowledge that, if present, income targeting behaviour may incentivise subjects to produce more, forgoing 'quality'. Once these considerations are accounted for, evidence from piece rates designs suggests that the morale effect of pay inequality dominates the incentive for income targeting.

Overall, despite an initial hope to obtain a more transparent picture of the effect of pay inequality on effort at this point, evidence on this hypothesis remains yet mixed and inconclusive.

Hence, further research is needed, especially in the form of appropriately designed experiments. With this objective in mind, the next section draws from the key findings of this paper to provide a set of recommendations to the design of laboratory experiments in the future.

5.2 Recommendations for future research

We conclude this discussion by delineating a set of recommendations for future laboratory designs. These recommendations will serve two purposes: i) to improve the control of potential confounds, such as reciprocity responses or income targeting incentives; and ii) to isolate the effect of pay inequality on morale and effort stemming from horizontal fairness considerations.

In general, experimental designs should:

- *Consider procedural fairness:* it is important for experiments to control for employees' perception of fairness in relation to pay inequality, in particular, whether it is perceived to be justified or unjustified. While there can be several ways to do this, the approach of Bracha et al. (2015) and Cardella and Roomets (2022), who anchor subjects' beliefs of their ability using an assessment task at the start of the experiment, and then induce pay inequality via a random allocation mechanism, seems to be a promising method to generate unjustified pay inequality.

- *Choose the appropriate control / testing approach:* while finding evidence in support of hypothesis H-C does not allow to conclude anything about the effect of pay inequality on effort—lower-paid employees will naturally exert less effort than higher-paid employees, especially under piece-rate payment schemes—it has been established that both {private; inequality} and {public; equality} treatments can serve as appropriate controls. However, it would be interesting to design experiments implementing both treatments, and to investigate which of these would lead to higher effort outcomes. Doing so would help to understand whether it is pay secrecy (private inequality), or wage compression (public equality), the most effective way to attenuate any existing effect of pay inequality on employees' effort.

- *Control for loss aversion:* which could explain the relative stronger response of effort to disadvantageous pay inequality treatments in the existing literature. Further, loss aversion is a key psychological driver of the incentive to increase effort when subjects perceive to be underpaid under piece rates. In particular, in piece-rate designs using

{private; inequality} as the control treatment (hypothesis H1-A), it could be useful to measure subjects' loss aversion, and then use this information as an additional control when analysing experimental data. While there is no widely accepted standard method to measure loss aversion that is applicable to riskless environments, a starting point would be to follow the approach taken by Gächter, Johnson, and Herrmann (2022) and/or Fehr and Goette (2007).

In three-person gift-exchange games, designs should aim to control for reciprocity behaviour (vertical fairness) and reputation building:

- *Make the employer's choice set common knowledge*: which will enable to control for the effect of vertical fairness considerations even in control treatments in which information on relative pay is private or there is pay equality. As noted, a number of studies may have achieved this by eliciting employees' choices through the strategy method (Gächter et al., 2012; Nosenzo, 2013)—according to which employees sequentially become aware of all the possible wages that they can be paid. However, the strategy method cannot be used in experiments featuring a real effort task.

- *Run an additional treatment without the employer choosing pay:* which should eliminate vertical fairness considerations (see, for instance, Study 3 of Gächter and Thöni (2010) in which the subject-employer becomes passive). Hence, this paper recommends the design of experiments featuring {exogenous; flat wage} treatments, in which wages are either manipulated by the experimenter (but subjects do not know this), or set by a random mechanism.

- *Avoid repeated interactions:* that could lead to incentives for reputation building, especially if the employees-employer match remains the same throughout the whole experiment. Repeated interactions could also introduce other confounds out of the control of the experimenter (e.g. learning effects).

When using piece rates, designs should aim to control for income targeting behaviour, which may generate an incentive to increase effort when subject-employees are paid the lower pay rate:

- *Induce an artificial income target:* when adopting the {private; inequality} treatment as the control treatment (hypothesis H1-A). As such, the income targeting incentive will be present even in the control treatment, and any evidence of treatment effects when

information on relative pay is public can be attributed, more exclusively, to the morale effect of pay inequality on effort. A possible approach would be to manipulate subjects' expectations about their final earnings as in Abeler, Falk, Goette, and Huffman (2011).

- *Set a large pay gap:* which would discourage income targeting behaviour by lower-paid employees. Doing so would then allow a more clean identification of the morale effect of pay inequality, independently of whether the design uses {private; inequality} or {public; equality} as control treatment.

- *Recognize the tradeoff between quantity and quality:* doing so would allow to identify more clearly the incentive given by piece rates (and the effect of income-targeting behaviour) on effort. This would also enable to get a better picture of the consequences of pay inequality on effort, when there is a trade-off between quantity and quality of output produced.

6 Concluding Remarks

The hypothesis that pay inequality affects worker effort finds support in behavioural economic theories suggesting that social comparisons have a major role in influencing employees' perception of fairness and morale in the workplace. The very existence of pay secrecy and pay compression practices is testimony to the fact that these concerns are common among compensation managers. However, despite a rich body of empirical literature, the evidence stemming from controlled laboratory experiments is mixed, and our understanding of the effect of pay inequality on effort still incomplete.

This paper contributed to this literature from the perspective of laboratory experiments. First, it has provided a novel taxonomy to classify existing laboratory designs investigating the effect of pay inequality on effort. These can be categorised between multi-lateral gift-exchange games and designs adopting piece-rate payments schemes. In so doing, the paper highlighted important methodological differences and similarities in the literature, as well as the existence of a number of confounds that may have impaired the identification of the effect of pay inequality on effort. Second, the paper developed a model of worker preferences that is suitable to characterise subjects' effort responses in these experiments. In fact, the model enables to analytically distinguish between the impact of horizontal pay comparison on worker effort, and other important fairness evaluations that could stem from reciprocity towards a subject-employer, or from concerns about income inequality—alongside those about pay inequality. A key contribution of

the paper consisted in adopting this conceptual framework to provide a formal assessment of the ability of these designs to identify the effect of horizontal pay comparison on effort.

This assessment has shown that failure to account for a number of confounds, such as reciprocity towards a subject-employer in multi-lateral gift-exchange games, and income targeting behaviour in designs adopting piece-rate payments—as well as other important design features such as the role of procedural fairness, and incentives for reputation building in repeated interactions—may lead to inaccurate interpretations of experimental data and evidence. In light of this assessment, this paper calls for a more cautious and informed interpretation of the evidence reported in the literature, and delineates a set of recommendations on how to improve the identification of the effect of pay inequality on effort for the development of laboratory designs in the future.

To conclude, the findings of this paper also bear a wider implication for the development of laboratory experiments in a multi-worker environment. In fact, the assessment of this paper suggests that whether experiments are aiming to identify the effect of either pay inequality, reciprocity, or income inequality, on workers' perceptions of fairness and effort, it is crucially important to ensure that the other two dimensions for fairness considerations are either appropriately controlled for, or removed, by design.

References

Abeler, J., S. Altmann, S. Kube, and M. Wibral (2010). Gift exchange and workers' fairness concerns: When equality is unfair. *Journal of the European Economic Association 8*(6), 1299 – 1324.

Abeler, J., A. Falk, L. Goette, and D. Huffman (2011). Reference Points and Effort Provision. *American Economic Review 101*(April), 470–492.

Akerlof, G. A. and J. L. Yellen (1990). The Fair Wage-Effort Hypothesis and Unemployment. *Quarterly Journal of Economics CV*(2), 255–283.

Altmann, S., A. Falk, A. Grunewald, and D. Huffman (2014). Contractual Incompleteness, Unemployment, and Labour Market Segmentation. *Review of Economic Studies 81*(October 2013), 30–56.

Angelova, V., W. Güth, and M. G. Kocher (2012). Co-employment of permanently and temporarily employed agents. *Labour Economics 19*(1), 48 – 58.

Bartling, B. and F. A. von Siemens (2011). Wage Inequality and Team Production: An Experimental Analysis. *Journal of Economic Psychology 32*(1), 1 – 16.

Bewley, T. F. (1999). *Why Wages Don't Fall During a Recession*. London: Harvard University Press.

Bewley, T. F. (2007). Fairness, Reciprocity, and Wage Rigidity. In P. Diamond and H. Vartiainen (Eds.), *Behavioural Economics and Its Applications*. Princeton University Press.

Blau, G. (1994). Testing the Effect of Level and Importance of Pay Referents on Pay Level Satisfaction. *Human Relations 47*(10), 1251–1268.

Bolton, G. E., J. Brandts, and A. Ockenfels (2005). Fair procedures: Evidence from games involving lotteries. *The Economic Journal 115*(506), 1054–1076.

Bolton, G. E. and A. Ockenfels (2000). A Theory of Equity, Reciprocity, and Competition. *American Economic Review 90*(1), 166–193.

Bracha, A., U. Gneezy, and G. Loewenstein (2015). Relative Pay and Labor Supply. *Journal of Labor Economics 33*(2), 297–315.

Brandts, J., J. M. Ortiz, and C. S. Belda (2019). Distributional concerns in managers' compensation schemes for heterogeneous workers: Experimental evidence. *Review of Behavioral Economics 6*(3), 193–218.

Breza, E., S. Kaur, and Y. Shamdasani (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics 133*(2), 611–663.

Burchett, R. and J. Willoughby (2004). Work productivity when knowledge of different reward systems varies: Report from an economic experiment. *Journal of Economic Psychology 25*(5), 591–600.

Cardella, E. and A. Roomets (2022). Pay distribution preferences and productivity effects: An experiment. *Journal of Behavioral and Experimental Economics 96*, 101814.

Charness, G., R. Cobo-Reyes, J. A. Lacomba, F. Lagos, and J. M. Perez (2016). Social comparisons in wage delegation: experimental evidence. *Experimental Economics 19*(2), 433–459.

Charness, G., U. Gneezy, and A. Henderson (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization 149*, 74 – 87.

Charness, G. and P. Kuhn (2007). Does Pay Inequality Affect Worker Effort? Experimental Evidence. *Journal of Labor Economics 25*(4), 693–723.

Cohn, A., E. Fehr, B. Herrmann, and F. Schneider (2014). Social Comparison and Effort Provision: Evidence from a Field Experiment. *Journal of the European Economic Association 12*, 877–898.

Cullen, Z. and R. Perez-Truglia (2022). How much does your boss make? the effects of salary comparisons. *Journal of Political Economy 130*(3), 766–822.

Dickson, A. and M. Fongoni (2019). Asymmetric reference-dependent reciprocity, downward wage rigidity, and the employment contract. *Journal of Economic Behavior & Organization 163*, 409 – 429.

Dufwenberg, M. and G. Kirchsteiger (2000). Reciprocity and wage undercutting. *European Economic Review 44*(4), 1069 – 1078.

Dufwenberg, M. and G. Kirchsteiger (2004, may). A Theory of Sequential Reciprocity. *Games and Economic Behavior 47*(2), 268–298.

Falk, A. and U. Fischbacher (2006, feb). A Theory of Reciprocity. *Games and Economic Behavior 54*(2), 293–315.

Fehr, E. and L. Goette (2007, March). Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment. *American Economic Review 97*(1), 298–317.

Fehr, E. and K. M. Schmidt (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics 114*(3), 817–868.

Flynn, J. (2022). Salary disclosure and individual effort: Evidence from the national hockey league. *Journal of Economic Behavior & Organization 202*, 471–497.

Frank, R. H. (1984). Are Workers Paid their Marginal Products? *The American Economic Review 74*(4), 549–571.

Gächter, S., E. J. Johnson, and A. Herrmann (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision 92*(3), 599–624.

Gächter, S., D. Nosenzo, and M. Sefton (2012). The impact of social comparisons on reciprocity. *The Scandinavian Journal of Economics 114*(4), 1346–1367.

Gächter, S. and C. Thöni (2010). Social Comparison and Performance: Experimental Evidence on the Fair Wage-effort Hypothesis. *Journal of Economic Behavior & Organization 76*(3), 531 – 543.

Gertler, M. and A. Trigari (2009). Unemployment Fluctuations with Staggered Nash Wage Bargaining Nash Wage Bargaining. *Journal of Political Economy 117*(1), 38–86.

Goette, L., D. Huffman, and E. Fehr (2004). Loss aversion and labor supply. *Journal of the European Economic Association 2*(2-3), 216–228.

Goodman, P. S. (1974). An Examination of Referents Used in the Evaluation of Pay. *Organizational Behavior and Human Performance 12*(2), 170 – 195.

Greiner, B., A. Ockenfels, and P. Werner (2011). Wage transparency and performance: A real-effort experiment. *Economics Letters 111*(3), 236 – 238.

Gross, T., C. Guo, and G. Charness (2015). Merit pay and wage compression with productivity differences and uncertainty. *Journal of Economic Behavior and Organization 117*, 233–247.

Güth, W., M. Königstein, J. Kovács, and E. Zala-Mezo (2001). Fairness within firms: The case of one principal and multiple agents. *Schmalenbach Business Review 53*(2), 82–101.

Hennig-Schmidt, H., A. Sadrieh, and B. Rockenbach (2010). In search of workers' real effort reciprocity—a field and a laboratory experiment. *Journal of the European Economic Association 8*(4), 817–837.

Hesse, N. and M. F. Rivas (2015). Does managerial compensation affect worker's effort? *Journal of Applied Economics 18*(2), 297–323.

Kahneman, D. (1992). Reference Points, Anchors, Norms, and Mixed Feelings. *Organizational Behaviour and Human Decision Processes* (51), 296–312.

Kahneman, D. and A. Tversky (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica 47*(2), 263–292.

Ku, H. and T. C. Salmon (2012). The incentive effects of inequality: An experimental investigation. *Southern Economic Journal 79*(1), 46–70.

Loewenstein, George F. Thompson, L. and M. H. Bazerman (1989). Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology 57*(3), 426–441.

Macera, R. and V. L. te Velde (2018). On the Power of Surprising versus Anticipated Gifts in the Workplace. *Working Paper*.

Maximiano, S., R. Sloof, and J. Sonnemans (2007). Gift exchange in a multi-worker firm. *The Economic Journal 117*(522), 1025–1050.

McDonald, I. M., N. Nikiforakis, N. Olekalns, and H. Sibly (2013). Social Comparisons and Reference Group Formation: Some Experimental Evidence. *Games and Economic Behavior 79*, 75–89.

Nosenzo, D. (2013). Pay secrecy and effort provision. *Economic Inquiry 51*(3), 1779–1794.

Ordóñez, L. D., T. Connoly, and R. Coughlan (2000). Multiple Reference Points in Satisfaction and Fairness Assessment. *Journal of Behavioural Decision Making 13*, 329–344.

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review 83*(5), 1281–1301.

Sliwka, D. and P. Werner (2017). Wage Increases and the Dynamics of Reciprocity. *Journal of Labor Economics 35*(2), 299–344.

Smith, J. C. (2015). Pay growth, Fairness and Job Satisfaction: Implications for Nominal and Real Wage Rigidity. *Scandinavian Journal of Economics 117*(3), 852–877.

Snell, A. and J. P. Thomas (2010). Labor Contracts, Equal Treatment, and Wage-Unemployment Dynamics. *American Economic Journal: Macroeconomics 2*(3), 98–127.

Sseruyange, J. and E. Bulte (2020). Wage differentials and workers' effort: Experimental evidence from uganda. *Oxford Bulletin of Economics and Statistics 82*(3), 647–668.

Tversky, A. and D. Kahneman (1991). Loss Aversion in Riskless Choice: a Reference Dependent Model. *Quarterly Journal of Economics 106*(4), 1039–1061.

A Appendix

A.1 Gift-exchange and reciprocity

In this section we show that the impact of morale stemming from vertical fairness considerations, as modelled by the function G in Section 3.2, is analogous to the "reciprocity payoff" characterising a more general model of a gift-exchange game with reciprocity. To do so, we build on the theory of sequential reciprocity developed by Dufwenberg and Kirchsteiger (2004), and we apply it to a two-player game setting with one-sided reciprocity. That is, we assume the worker is motivated by reciprocity, while the firm is motivated only by profit (see also Dufwenberg and Kirchsteiger (2000) and Dickson and Fongoni (2019)).

Consider a worker i and a firm f, playing a canonical, one-shot, gift-exchange game. The firm moves first and chooses the wage $w_f \in \{w^L, w^H\}$ that maximises their material payoff π_f, which is given by profit: $\pi_f = e_i - w_f$. The worker moves second, and chooses effort $e_i \in [0, e^{max}]$ that maximises their utility:

$$u_i = \pi_i + Y_i \kappa_{if} \chi_{i\,fi},$$

where $\pi_i = w_f - d(e_i)$ is the worker's material payoff (the function d is defined in Section 3.2), and $Y_i \kappa_{if} \chi_{ifi}$ is the worker's reciprocity payoff with respect to the firm f (Dufwenberg and Kirchsteiger, 2004): Y_i is a non-negative parameter describing i's sensitivity to reciprocity; κ_{if} is the worker i kindness to the firm f, it is positive if i is kind and negative if i is unkind; $\chi_{i\,fi}$ is the worker i's belief about how kind the firm f is to the worker i. It is positive if the worker i believes j is kind to them, and negative if the worker i believes j is unkind to them (mathematically $\chi_{i\,fi}$ is equivalent to κ_{fi}—i.e. how kind is f to i—although it captures a psychological component that pertains to i and not to f). If $Y_i > 0$ then: if i believes that f is kind to them (i.e. $\chi_{i\,fi} > 0$), then i's reciprocity payoff with respect to f is increasing in i's kindness to f. If i believes that f is unkind to them (i.e. $\chi_{i\,fi} < 0$), then i's reciprocity payoff with respect to f is decreasing in i's kindness to f. This is the way in which u_i reflects the idea that if i thinks that f is kind (unkind) to them, then i wants to be kind (unkind) in return. Another key feature is that κ_{if} and $\chi_{i\,fi}$ depend on worker i's beliefs. κ_{if} is measured by comparing the material payoff that i believes that f gets to the set of material payoffs that i believes that f would get were i to choose differently than he does. Effectively, i is kind if he

believes he gives f 'a lot' relative to what i believes he could give to f in principle. Conversely, i is unkind if he believes he gives f 'very little' relative to what i believes he could give to f in principle. $\chi_{i\,fi}$ is measured similarly, except one has to 'move up' a level in the belief hierarchy. For example, i believes f is kind if i believes f believes she gives i 'a lot' relative to what f believes she could give to i in principle.

An important ingredient of this class of models is the definition of kindness. Rabin (1993) and Dufwenberg and Kirchsteiger (2004) define i's kindness as the difference between what i believes he gives to f, π_f, and the average of the maximum and minimum payoff that i believes he could give to f in principle, which we denote by $\pi_f^{r_i}$. Hence, $\kappa_{if} = \pi_f - \pi_f^{r_i}$, where $\pi_f^{r_i}$ is a reference point measuring how kind is i to f (for a given i's belief of what f is going to do, normally denoted b_{if}). Similarly we define the worker's belief of f's kindness, as the difference between what f believes he gives to i, π_i, and the average of the maximum and minimum payoff that f believes he could give to i in principle, which we denote by $\pi_f^{r_f}$. Hence, $\chi_{i\,fi} = \pi_i - \pi_f^{r_f}$, where $\pi_f^{r_f}$ is a reference point measuring how kind is f to i (for a given i's belief of f's belief of what player i is going to do, normally denoted by $c_{i\,fi}$). Nevertheless, as also Dufwenberg and Kirchsteiger (2004) point out, there is no deep justification for picking the average of the maximum and minimum payoffs, except that the choice is simple and does not affect the qualitative performance of the theory.

First note that without reciprocity, the worker will choose $e_i = \bar{e}_i > 0$ ($\bar{e}_i$ defined in Section 3.2), for any given w_f and the firm will choose $w_f = w^L$. Next, to study the game under reciprocity it is necessary to characterise κ_{if} and $\chi_{i\,fi}$ (noting that in this setting we do not model the firm's decision, and the worker does not have to form beliefs about what wage the firm has chosen, since they actually observe the wage before choosing effort). Let us first characterise $\chi_{i\,fi}$, the belief of the worker about how kind the firm has been with them. By definition, $\chi_{i\,fi} = \pi_i - \pi_f^{r_f}$ where $\pi_i = w_f - d(\tilde{e}_i)$ and $\pi_f^{r_f} = \frac{1}{2}[w^H - d(\tilde{e}_i) + w^L - d(\tilde{e}_i)]$, where $\tilde{e}_i$ captures worker i belief of firm f's belief of how much effort the worker will exert. This implies that

$$\chi_{i\,fi} = w_f - \frac{1}{2}[w^H + w^L].$$

Then, we can characterise $\kappa_{if} = \pi_f - \pi_f^{r_i}$, which is a measure of how kind the worker is to the firm. By definition, $\kappa_{if} = \pi_f - \pi_f^{r_i}$ where: $\pi_f = e_i - w_f$ and $\pi_f^{r_i} = \frac{1}{2}[e^{max} - w_f + \bar{e}_i - w_f]$, for any given w_f chosen by the firm, and effort level e_i chosen by the worker (note $e_i = 0$ is never

Pareto efficient for the worker given our assumption on d). This implies that

$$\kappa_{if} = e_i - \frac{1}{2}[e^{max} + \bar{e}_i].$$

Hence, for any given wage offer w_f we can rewrite the worker's utility as:

$$u_i = w_f - d(e_i) + Y_i \left\{ e_i - \frac{1}{2}[e^{max} + \bar{e}_i] \right\} \left\{ w_f - \frac{1}{2}[w^H + w^L] \right\}. \tag{11}$$

The worker believes the firm has been kind to them if $w_f > \frac{1}{2}[w^H + w^L]$ and unkind if $w_f < \frac{1}{2}[w^H + w^L]$, hence we can define $r_i^G \equiv \frac{1}{2}[w^H + w^L]$ as the worker's reference wage for vertical fairness evaluations. Analogously, we can define $e_f^R \equiv \frac{1}{2}[e^{max} + \bar{e}_i]$ as a reference effort level to measure the worker kindness towards the firm. In the paper, we assumed for simplicity that $r_i^G = w^H$ (the reference for the firm's kindness is given by their highest available alternative) and that $e_f^R = \bar{e}_i$ (the reference for the worker's kindness is given by the effort the worker would exert absent reciprocity considerations). Finally, introducing asymmetries in fairness evaluations (which in the model of Section 3.2 are justified by loss aversion), that is, $Y_i = 1$ when $w_f > r_i^G$ and $Y_i = \lambda > 1$ when $w_f < r_i^G$, it turns out that the utility in (11) is analogous to (1) in which $\mathbf{1}^G = 1$, and $\mathbf{1}^S = \mathbf{1}^I = 0$ as required.

A.2 Piece rates and income targeting

In Section 4.2 we assumed that if the information on the co-worker piece rate z_j is known, then worker i will expect ω_j to be higher (lower) than their wage, ω_i, if worker j is paid a higher (lower) piece rate. That is, we assumed that $\mathbb{E}\omega_j \gtreqless \omega_i$ if $z_j \gtreqless z_i$. In this section we will show that this assumption holds if the pay gap $|z_i - z_j|$ is sufficiently large.

First of all note that $\mathbb{E}\omega_j = \mathbb{E}[e_j z_j] = \mathbb{E}e_j z_j$, since z_j is known. Hence, whether the assumption holds crucially depends on the worker i's expectations about of how much effort their co-worker j will exert, denoted by $\mathbb{E}e_j$. Nevertheless, the behaviour of j also depends on their expectations about how much effort worker i will exert, denoted by $\mathbb{E}e_i$. Thus, worker i choice of effort will depend on their expectations about their co-worker j's expectations about their (worker i's) effort. An analogous reasoning applies to worker j. We are therefore interested in situations in which these expectations are consistent with each other for given z_i and z_j.

Let us consider the case in which $z_i < z_j$, which is also the one considered in the assessment of Section 4.2. There are three possible scenarios: i) worker i expects that worker j's effort is

such that $e_i z_i = \mathbb{E}e_j z_j$ when $z_i < z_j$, and worker j expects that worker i's effort is such that $\mathbb{E}e_i z_i = e_j z_j$; ii) worker i expects that worker j's effort is such that $e_i z_i > \mathbb{E}e_j z_j$ when $z_i < z_j$, and analogously worker j expects $\mathbb{E}e_j z_i < e_j z_j$; iii) worker i expects that worker j's effort is such that $e_i z_i < \mathbb{E}e_j z_j$ when $z_i < z_j$, and analogously worker j expects $\mathbb{E}e_i z_i > e_j z_j$.

For scenario i) to be consistent with workers' expectations, it must be that $\mathbb{E}e_i z_i = \mathbb{E}e_j z_j$ when $z_i < z_j$; for scenario ii) it must be that $\mathbb{E}e_i z_i > \mathbb{E}e_j z_j$ when $z_i < z_j$; and for scenario iii) it must be that $\mathbb{E}e_i z_i < \mathbb{E}e_j z_j$ when $z_i < z_j$. Let us now consider each scenario in turn.

Scenario i). This case would imply that workers expects their final earnings to be the same, hence $\omega_i = r_i^I$ and $\omega_j = r_j^I$, which implies that:

$$\mathbb{E}e_i = \bar{e} + \alpha\lambda[z_i - z_j] + [1 - \gamma]z_i$$
$$\mathbb{E}e_j = \bar{e} + \alpha[z_j - z_i] + [1 - \gamma]z_j.$$

For this scenario to be consistent with workers' expectations, it must be that $\mathbb{E}e_i z_i = \mathbb{E}e_j z_j$, hence it must be that:

$$\bar{e}z_i + \alpha\lambda[z_i - z_j]z_i + [1 - \gamma]z_i^2 = \bar{e}z_j + \alpha[z_j - z_i]z_j + [1 - \gamma]z_j^2.$$

Rearranging this equation, after some algebra, yields:

$$[1 - \gamma][z_j + z_i] = -\bar{e} - \alpha[z_j + \lambda z_i],$$

which never holds since the right-hand side is negative, while the left-hand side is positive. Hence, in this scenario workers' expectations are inconsistent with the outcome of their expected behaviour.

Scenario ii). This case would imply that worker i expects their final earnings to be higher than worker j and worker j expects their final earnings to be lower than worker i, hence $\omega_i > r_i^I$ and $\omega_j < r_j^I$, which implies that:

$$\mathbb{E}e_i = \bar{e} + \alpha\lambda[z_i - z_j] + z_i$$
$$\mathbb{E}e_j = \bar{e} + \alpha[z_j - z_i] + [1 - \gamma + \lambda\gamma]z_j.$$

For this scenario to be consistent with workers' expectations, it must be that $\mathbb{E}e_i z_i > \mathbb{E}e_j z_j$,

hence it must be that:

$$\bar{e}z_i + \alpha\lambda[z_i - z_j]z_i + z_i^2 > \bar{e}z_j + \alpha[z_j - z_i]z_j + [1 - \gamma + \lambda\gamma]z_j^2.$$

Rearranging this equation, after some algebra, yields:

$$\frac{z_j^2\gamma[\lambda - 1]}{z_j - z_i} < -\bar{e} - \alpha[z_j + \lambda z_i] - [z_j - z_i],$$

which, again, never holds since the right-hand side is negative, while the left-hand side is positive. Hence, in this scenario workers' expectations are inconsistent with the outcome of their expected behaviour.

Scenario iii). This case would imply that worker i expects their final earnings to be lower than worker j and worker j expects their final earnings to be higher than worker i, hence $\omega_i < r_i^I$ and $\omega_j > r_j^I$, which implies that:

$$\mathbb{E}e_i = \bar{e} + \alpha\lambda[z_i - z_j] + [1 - \gamma + \lambda\gamma]z_i$$
$$\mathbb{E}e_j = \bar{e} + \alpha[z_j - z_i] + z_j.$$

For this scenario to be consistent with workers' expectations, it must be that $\mathbb{E}e_i z_i < \mathbb{E}e_j z_j$, hence it must be that:

$$\bar{e}z_i + \alpha\lambda[z_i - z_j]z_i + [1 - \gamma + \lambda\gamma]z_i^2 < \bar{e}z_j + \alpha[z_j - z_i]z_j + z_j^2.$$

Rearranging this equation, after some algebra, yields:

$$\frac{z_i^2\gamma[\lambda - 1]}{z_j - z_i} < \bar{e} + \alpha[z_j + \lambda z_i] + [z_j - z_i],$$

which may hold for a given configuration of parameters. In fact, it can be deduced that as z_j increases the right-hand side gets larger, while the left-hand side gets smaller. Hence, there exists a z_j such that the inequality above always holds, for given z_i, α, γ and λ. In this scenario workers' expectations can be consistent with the outcome of their expected behaviour, and we assume that the z_j is large enough so that this assumption holds.