Strategic Shirking in Competitive Labor Markets:

A General Model of Multi-Task Promotion Tournaments with Employer Learning*

Jed DeVaro

California State University, East Bay

(E-mail: jed.devaro@csueastbay.edu)

Oliver Gürtler

University of Cologne

(E-mail: oliver.guertler@uni-koeln.de)

13 October 2019

Abstract

In a multi-task, market-based promotion tournament model, under different environments concerning employer learning about worker ability, it is shown that:

i) Asymmetric learning in multi-task jobs is a necessary condition for "strategic shirking" (i.e., underperforming on certain tasks to increase the promotion probability).

ii) When learning becomes increasingly symmetric on one task, the effort allocated to that task could increase or decrease, but effort on the other task increases.

iii) Strategic shirking does not occur in equilibrium in single-task models.

iv) Promotions signal worker ability even when there is symmetric learning on one task, if learning is asymmetric on another.

Keywords: strategic shirking, promotions, classic and market-based tournaments, symmetric learning, asymmetric learning, inefficient job assignments, inefficient effort allocation

JEL classification: J24, M53

*We thank the editor (Ramon Casadesus-Masanell), an anonymous coeditor and two anonymous referees, Hugh Cassidy, Antti Kauhanen, Fabio Mariani, Mike Waldman, and seminar participants at University of New South Wales, University of Sydney, California State University East Bay, National University of Singapore, Linnaeus University, Humboldt University Berlin, the 2017 cbess conference on Contests: Theory and Evidence, the 92nd Annual Conference of the Western Economic Association International, and the Southern Economic Association 87th Annual Meetings, for helpful comments. We are particularly grateful to Hugh Cassidy, whose input contributed significantly to the material on overlapping wage distribution across job levels.

1 Introduction

Ambitious, career-conscious workers are ever alert to ways they can outshine their peers in promotion contests. Talent and hard work are well-known ingredients for success, but their roles in determining promotions complexify when jobs are comprised of multiple tasks, as is usually the case in organizations. For example, police officers engage in primary tasks (e.g., making traffic stops, and pursuing and apprehending suspects) and, to a lesser extent, in managerial and leadership tasks, and public speaking. For police chiefs, however, the relative emphases placed on those activities are reversed. Similarly, school teachers are primarily engaged with educating students, whereas managerial and leadership activities (beyond the classroom setting) are not central to their job. For the school principals who supervise them, however, the relative emphases on those tasks are reversed. In both settings, the police officers and teachers who do stellar work on their primary tasks might actually be denied promotions because they are deemed indispensable in their current jobs, relative to their coworkers. A better strategy for achieving promotion to police chief, or to principal, might be to purposely underperform (or "strategically shirk") on the primary tasks while overperforming on the leadership tasks that are emphasized in the managerial job.[1]

Strategic shirking is a fundamental incentive problem arising in multi-task promotion tournaments, but it has only been analyzed in "classic" settings in which the employer pre-commits to the wages paid to promoted and non-promoted workers to elicit desired worker behavior (e.g., effort choices) without direct influence from an outside market of competing firms. The goal of this study is to analyze the problem in "market-based" settings in which wages are determined by the bids of competing firms, as in Waldman (1984). This contribution in a multi-task setting parallels recent developments in the (single-task) tournament literature. The bulk of that literature, starting with Lazear and Rosen (1981), concerns "classic tournaments" in which the employer can pre-commit to the prize structure. Recently, a literature on (single-task) "market-based tournaments" has emerged, based on the realistic idea that the bids of competing employers make pre-commitment challenging, if not impossible.[2]

Moving from a classic to a market-based wage-setting mechanism is considerably more challenging in the multi-task environment than in the single-task environment that the previous literature has addressed. One reason is that the nature of the process by which competing employers learn about workers' talents becomes more complicated when there are multiple tasks. Learning is crucial in a market-based setting, because external compensation offers hinge on what information potential employers learn about a worker's abilities, which will often be less than the information available to the worker's current employer. When jobs are comprised of multiple tasks, however, many possible learning environments arise. Competing employers might learn about primary-task performance and leadership-task performance at the same rate as the worker's

[1] Henceforth, the term "primary tasks" will be used to refer to the activities on which workers (such as police officers and teachers) should be focused, and "leadership tasks" will be used to refer to the activities that are emphasized more in the supervisory positions.

[2] See, for example, Gibbs (1995), Zábojník and Bernhardt (2001), Waldman (2013), Gürtler and Gürtler (2015), and DeVaro and Kauhanen (2016).

current employer, as in the case of pure symmetric learning. Alternatively, the current employer might learn performance on both tasks, whereas competing employers only learn that worker's job assignment, as in the case of pure asymmetric learning. Learning might also be symmetric on one task and asymmetric on the other. Many intermediate cases are possible, involving mixtures of symmetric and asymmetric learning, in which either or both tasks are observed by competing employers with certain probabilities.

The forthcoming analysis accommodates a range of possible learning environments. The theoretical model is quite general, integrating the core features from the two pillars of the theoretical promotions literature (i.e., tournament models and learning-based job-assignment models), both of which are heavily focused on single-task jobs. Most models in the promotions literature can be classified based on the following typology of features:

1. Fixed versus flexible managerial job slots[3]
2. Single-task versus multi-task jobs
3. Endogenous worker choices (e.g., effort) versus not
4. Heterogeneous versus homogeneous worker ability
5. Symmetric versus asymmetric employer learning about worker ability[4]
6. Single firm versus multiple firms
7. Classic versus market-based wage setting

The present model fully incorporates all of these dichotomies, assuming fixed managerial job slots in the main analysis, with flexible slots covered in an extension. The model nests learning that is purely asymmetric, purely symmetric, or a blend of both types, and it emphasizes multi-task jobs while considering single-task jobs in an extension. Both worker ability and endogenous effort choices are incorporated and are task-specific. Although the main focus is on market-based wage setting with multiple competing firms, another extension addresses classic wage setting in a single-firm context.

The model is the first in the literature to analyze task-specific efforts and abilities in multi-task job hierarchies in competitive labor markets, under a range of assumptions about employer learning. Such a general model contributes in three ways. First, it allows results on strategic shirking, which have recently been developed in the context of single-firm models in DeVaro and Gürtler (2016a,b), to be extended to more realistic and complex environments in which multiple employers compete in the labor market under various assumptions about employer learning. Second, it generates new insights, beyond simply the sum of those implied by both pillars of the theoretical promotions literature considered separately. Third, it provides an organizing framework for understanding the literature and how various theoretical results relate to particular

[3]Fixed slots mean that the number of managerial positions is limited (usually to one) so that workers compete with each other for promotion, whereas flexible slots mean that everyone with sufficiently high performance gets promoted.

[4]Employer learning about worker abilities is typically neglected in tournament models, because the absence of post-promotion production renders the issue of efficient job assignment irrelevant, whereas employer learning is central to job-assignment models and can be either symmetric or asymmetric. Examples of symmetric learning models of promotion include Gibbons and Waldman (1999, 2006), DeVaro and Morita (2013), and Smeets et al. (2019). Examples of asymmetric learning models of promotion include Waldman (1984), Bernhardt (1995), Ghosh and Waldman (2010), DeVaro and Waldman (2012), and Cassidy et al. (2016).

modeling assumptions.

Concerning the first of these contributions, a key difference between a single-firm environment and a multi-firm environment is information revelation, i.e., how much information about worker ability do competing firms possess (versus incumbent firms) and what assumptions govern how further information is revealed to competing firms over time? Such considerations involving information allocation and revelation are absent in the single-firm case and drive the competitive wage bidding in the multiple-firm case. When there are multiple firms, wage determination involves a process of competitive bidding, which is more realistic than the single-firm case in which the incumbent employer chooses the wage.

Consider a labor market of competing employers, each of which has a two-level hierarchy involving subordinate jobs and managerial jobs. Suppose that two tasks (e.g., primary and leadership) are used in each of the two jobs, though with primary tasks emphasized relatively more in the subordinate job's production function. Those subordinates who perform sufficiently well are promoted to managers, where they enjoy higher pay, whereas the others remain as subordinates. Two questions arise concerning equilibrium behavior. First, do subordinates strategically shirk on primary tasks? Second, do they exert more effort, or less, when the learning environment becomes increasingly symmetric and less asymmetric?

The answer to the first question is yes. Subordinates strategically shirk on primary tasks while hustling to perform exceptionally well on leadership tasks. They engage in such distortion of effort across the two tasks because of their employer's inclination to promote the worker who has demonstrated the potential to be more productive as a manager than as a subordinate. Interestingly, the nature of employer learning is shown to matter for the preceding result. In particular, if learning is purely symmetric (i.e., symmetric on both tasks), then strategic shirking does not occur on either task. Thus, asymmetric learning is a necessary, but not sufficient, condition for strategic shirking in multi-task, market-based promotion models.

Concerning the second question, the degree to which effort allocation is inefficient depends on the employer learning environment. The reason is that workers wish to impress potential employers if those employers are able to observe workers' performance, but also that the learning environment has implications for wages, and wage differences between job levels create incentives to strategically shirk. When the learning about ability on one task becomes increasingly symmetric, the effect on that task's effort is ambiguous, but effort on the other task increases. It is possible that efficiency in effort allocation may decline as learning about ability becomes increasingly symmetric.

A policy implication is that employer practices designed to increase effort and that are frequently coupled with promotion-based incentives (e.g., incentive compensation) might exacerbate an inefficiency in effort allocation, for example by inducing workers to overwork on certain tasks. In the examples from the beginning of this section, incentive compensation could be detrimental on leadership tasks.[5] Moreover, this result overturns the natural intuition that labor market policies, technological change, or other institutional forces

[5]In a single-firm setting, devoid of labor-market competition, DeVaro and Gürtler (2016a) find that inefficiencies from strategic shirking in a promotion-based incentive system with classic wage-setting can be mitigated if incentive compensation is used on the task that is emphasized in the subordinate job and on which workers strategically shirk.

that move the labor market further from asymmetric learning and closer to symmetric learning must increase welfare.

Whereas the preceding results concern inefficiency in effort allocation, the job-assignment literature focuses instead on inefficiency in job assignments. A celebrated result from that literature is that job assignments (in a single-task setting) are inefficient when employer learning about worker ability is asymmetric, and efficient when employer learning is symmetric.[6] The main model of the present analysis assumes a fixed managerial job slot that must be staffed; under those assumptions, job assignments are fully efficient. In an extension in which managerial job slots are flexible, however, it is shown that the result concerning inefficient job assignments from the prior literature extends to a more general, multi-task setting with multi-dimensional abilities and efforts, as long as there is asymmetric learning on at least one task.

Following the presentation of the model and its extensions, some empirical implications are highlighted, the most interesting of which concerns overlapping wage distributions across adjacent job levels, which is an important phenomenon that prior, influential theoretical models have been unable to fully explain.

2 Background and related literature

Four prior studies provide empirical support for fixed job slots, as are assumed in the managerial job in the main analysis. In a pair of related studies using samples of employers from four U.S. metropolitan areas, DeVaro (2006a,b) provide evidence that relative performance determines promotions, which is a direct implication of fixed managerial slots in a job hierarchy. In other words, when a worker's competitors perform better, that worker's promotion chances suffer. In both of those studies, a worker's promotion probability is found to be increasing in that worker's employer-reported subjective performance rating and decreasing in the employer-reported subjective performance rating of the "typical worker" in that worker's same job (which is taken as a proxy for the performance of those against whom that worker competes for promotion) when both performance variables are simultaneously included in the statistical model. DeVaro and Kauhanen (2016) find the same result using matched, employer-employee panel data from Finland. In contrast to the earlier two papers, DeVaro and Kauhanen (2016) do not rely on employer-reported subjective performance ratings but rather derive inferred performance measures (both for the worker in question and for all other workers against whom that worker competes for promotion) from data on performance-based pay. Focusing on hiring rather than promotions, Lazear et al. (2018) extensively discuss fixed slots and their empirical implications. They also use four data sets to provide an array of empirical results consistent with fixed slots. First among these, and echoing the prediction from the three preceding studies of promotions, the authors

[6] The result is due to Waldman (1984) and is based on the following logic. When a worker's employer has private information about the worker's ability, promoting (or not promoting) that worker signals the worker's ability to competing employers, who update their wage bids accordingly. The initial employer then must match competing wage offers to retain the worker. Thus, some workers who would be marginally more productive if promoted may be inefficiently denied promotions so that their employers can avoid the resulting costly wage increases.

The findings by Dato et al. (2016), who conduct an experimental study to test Waldman's model, strongly support his results.

find in the oDesk data that the hiring probability depends not only on an applicant's skills but also on the skills of the competition.[7]

Tournament models assume fixed managerial job slots and allow for endogenous worker choices (typically effort levels but sometimes human capital investments or even perverse, unproductive behaviors such as sabotaging one's coworkers).[8] Although most tournament models are "classic", i.e., assume a single firm that can pre-commit to wage offers that are paid after workers exert effort, recently a market-based tournament literature has emerged in which multiple employers compete in the labor market.[9] When wage setting is classic, the single employer sets wages at each job level to elicit the desired worker behaviors. When wage setting is market-based, the wages at each job level are determined by the wage offers from competing employers.

Job-assignment models assume that the returns to ability are greater in higher-level jobs, so that it is efficient to promote the new hires who demonstrate high ability in entry-level jobs. Recently, such models typically feature a market of competing employers, though early job-assignment models (e.g., MacDonald 1982) ignore the possibility of workers moving between firms after the initial entry decision. They usually assume flexible managerial job slots and do not allow for endogenous worker choices, though there are exceptions. For example, Ghosh and Waldman (2010) assume flexible managerial job slots but allow for endogenous worker effort choices, Waldman and Zax (2019) assume flexible managerial job slots but allow for endogenous choices concerning human capital acquisition, and DeVaro and Morita (2013) do not allow for endogenous worker choices but assume fixed managerial job slots. The present model incorporates the key features of, e.g., Ghosh and Waldman (2010), while additionally incorporating multiple tasks, multi-dimensional effort and ability, a continuum of types of employer learning, and a managerial slot constraint.

Multi-task promotion models are rare in either pillar of the literature, though some recent work has emerged.[10] DeVaro and Gürtler (2016b) introduce the concept of strategic shirking in a job-assignment model without a hierarchy, and DeVaro and Gürtler (2016a) embed the mechanism into a job hierarchy. Each job in the two-level hierarchy has two tasks, and worker efforts and abilities are both task-specific. In both studies, workers strategically shirk on their least preferred task and work extra hard on their preferred

[7]The authors provide four further pieces of evidence that are consistent with fixed slots based on their theoretical framework. First, they find evidence of "bumping", which happens when applicants who are better qualified than the rest of the applicant pool lose out on their preferred job because they are "bumped" by someone who is even better. Some jobs end up getting filled by "overqualified" applicants (specifically those who were bumped) and others by "underqualified" applicants. Second, less able workers are more likely to be unemployed than high-ability workers because the latter are more versatile and able to perform well in a broad range of jobs. Third, high-skilled jobs have higher vacancy rates than low-skilled jobs. Fourth, the variance of pay rises over time for higher-skilled workers.

[8]The original work is Lazear and Rosen (1981), and more recent theoretical studies include Waldman (2003), Tsoulouhas et al. (2007), Gürtler and Kräkel (2010), Schöttner and Thiele (2010), Zábojník (2012), Waldman (2013), and Brown and Minor (2014). Representative empirical studies include Audas et al. (2004), DeVaro (2006a,b), Agrawal et al. (2006), Coffey and Maloney (2010), Brown (2011), and DeVaro and Kauhanen (2016). Chowdhury and Gürtler (2015) survey the literature on sabotage in tournaments.

[9]See, for example, Gibbs (1995), Zábojník and Bernhardt (2001), Waldman (2013), Gürtler and Gürtler (2015), and DeVaro and Kauhanen (2016).

[10]Waldman (2016) considers two types of ability (academic and productive). DeVaro and Gürtler (2016a,b) consider task-specific ability in a model with two tasks. Brilon (2015) considers a model with two different activities, each of which uses two different skills (e.g., technical and analytical). Prasad and Tran (2013) consider worker investments in two types of skills (general and firm-specific). In related work, Waldman and Zax (2019) consider the mix between investments in firm-specific and general human capital.

task. In the (2016b) study, workers' preferences over tasks are exogenously given, whereas in the (2016a) study, worker preferences are determined by the task that is emphasized in the promotion rule (and that is, therefore, associated with larger wage increases). Both studies are distinguished from the present work in that they assume a single firm, a classic wage-setting mechanism in which the employer can pre-commit to wage offers, and no role for employer learning in the labor market.

A result of the (2016b) study is that employers can mitigate strategic shirking by requiring greater balance in future job assignments (i.e., requiring workers to engage substantially in both tasks). Similarly, in the (2016a) study employers can combat strategic shirking by using a promotion rule that rewards greater balance in the task-specific performances. These solutions to strategic shirking require that firms can commit to future job assignments, which may be challenging in practice and which is disallowed in the present analysis, though discussed at the end of section 5. Benson et al. (2019) provide empirical evidence from the sales industry suggesting that, in fact, firms can commit to promotion rules that reward sales performance, even though other skills, such as leadership, are more valuable in managerial positions. Such employer behavior can be interpreted as an attempt to mitigate the strategic shirking problem.

The present study shows that although strategic shirking extends to a multiple-firm context with market-based wage setting, the mechanism driving that behavior becomes more complicated in a market-based setting. The reason is that the wage-generating process, which drives workers' desires to strategically shirk, hinges on what information competing firms have (relative to the worker's current employer) about worker ability, and with multi-dimensional ability there are many possibilities for employer learning. The analysis shows that the nature of the learning has implications for whether strategic shirking occurs; in particular, asymmetric learning is necessary but not sufficient for strategic shirking.

The present model features a tension between current incentives and future job assignments that evokes similar themes from Baker et al. (1988), Milgrom and Roberts (1988), Waldman (2003), Ke et al. (2017), and DeVaro and Gürtler (2016a,b). It also shares some features in common with Holmström's (1999) career concerns model. That model and the present one assume multiple firms and a single worker; in both models firms cannot commit to certain future wage payments (as is possible in classic tournament models); and in both models the worker's ability is initially unknown, the firms form beliefs on the worker's efforts that are used to form beliefs on ability given realized output, and the worker chooses effort to signal ability. The key differences are that in the present analysis there is a job hierarchy, effort choices signal ability on two dimensions rather than one, workers signal low ability on the task that is less emphasized in job 2, and there is no noise in the production function, conditional on ability.

The analysis relates to Milgrom and Oster (1987) in that both studies allow workers to vary in the extent to which their abilities are observable by competing employers in the market, both studies feature promotions that reveal information about (some) workers' abilities to competing employers, and both studies allow the incumbent firm to make a counteroffer in an attempt to retain the worker during an attempted raid. An important difference is that the present analysis incorporates multiple tasks, whereas Milgrom and

Oster (1987) assume a single task. A further difference concerns the assumptions governing how information about worker ability is revealed to the market. In Milgrom and Oster (1987) there are two types of workers. "Visibles" are those whose abilities are observable to all employers in the market, before and after promotion, so promotions offer no new information about ability. "Invisibles" are workers whose abilities cannot be observed by competing employers. An Invisible's own employer can observe the worker's ability after a period of production, using that information to make promotion decisions. Competing employers, however, while observing these workers' job assignments, only observe these workers' abilities if they are promoted, which creates the possibility that employers might "hide" their high-ability Invisibles to prevent them from being poached by competitors. In contrast, the present model assumes that workers' task-specific performances are revealed to competing employers only with certain probabilities, so that the nature of employer learning may differ across tasks. A further difference between the models, as noted in footnote 20, is that a strong winner's curse result emerges in Milgrom and Oster (1987), whereas the present analysis makes assumptions to eliminate the winner's curse.

The model also relates to Waldman and Zax (2019), which builds on the insight from Waldman (1984) that the inefficiency in job assignments that arises in promotion signaling models is mitigated when firm-specific human capital is important. The Waldman and Zax (2019) innovation treats human capital as an investment by either the worker or the employer. In particular, the party making the investment chooses between no human capital or a fixed positive amount, and in the latter case also chooses the mix of human capital between general versus firm-specific. Like the present study, that analysis allows competing firms to vary in what they observe about workers beyond simply their job assignments. Specifically, the equilibrium promotion policy and human capital acquisition are compared when human capital acquisition is observable and not observable by prospective employers. Under most scenarios, the result is that human capital investments are biased towards the accumulation of more firm-specific human capital, thereby mitigating the inefficiency in job assigments that arises from employers' asymmetric learning about workers' abilities. The analysis differs from the present one in that investments are in the mix of two types of human capital (as opposed to the level of effort), one task is considered rather than two, and flexible managerial job slots are assumed rather than fixed slots as in the present study's tournament setup.

3 A general model of careers in competitive labor markets

In a two-period setup with no discounting, consider a competitive labor market with $N \geq 3$ identical firms and an even number $n(< 2N)$ of workers; all parties are risk-neutral.[11] Each firm has a two-level job hierarchy with job 1 at the low level and job 2 at the high level. Each job consists of two tasks, α and β, that differ in importance across jobs. Explicit incentive schemes that link pay to performance are infeasible, as are long-term contracts that bind workers to a firm for both periods.

[11] Three firms is the minimum number that guarantees that there is always competition. Section 5.3 considers a number less than three.

Assume that at the start of period 1, all parties are price-takers, and expected worker productivity is highest if a firm hires exactly two workers. These assumptions ensure that, in equilibrium, firms hire two workers or none at all. For now, therefore, period-1 wages are taken as given and a representative firm is considered that has hired two representative workers at the start of period 1. The matching of workers and the determination of period-1 wages are considered in subsection 4.1.

Let t, k, j, h, and i index periods, jobs, tasks, firms, and workers, respectively. Let a_{ij} denote worker i's ability on task j, which at the outset is unknown to all parties (including worker i himself). Suppose a_{ij} is a random variable that is identically and independently distributed on $[\underline{a}, \overline{a}]$ $(0 < \underline{a} < \overline{a})$ with pdf f (that has full support) and cdf F.[12] The distribution is assumed to be symmetric around the (finite) mean, $E[a]$.

Let $e_{ij}^t \in \mathbb{R}$ denote worker i's effort on task j in period t, which is unobservable to the firms and to the other worker. As in DeVaro and Gürtler (2016a,b), the worker's effort cost in period t is $\kappa_i^t = \sum_j \kappa\left(e_{ij}^t\right)$ with $\kappa\left(e_{ij}^t\right) = \begin{cases} \frac{\kappa_-}{2}\left(e_{ij}^t\right)^2 & \text{for } e_{ij}^t \leq 0 \\ \frac{\kappa_+}{2}\left(e_{ij}^t\right)^2 & \text{for } e_{ij}^t > 0 \end{cases}$, $\kappa_+, \kappa_- > 0$. Over the region $e_{ij}^t \geq 0$, this cost function is standard in the literature, but over the region $e_{ij}^t < 0$ it requires discussion. Recall, first, that $e_{ij}^t = 0$ should not be taken literally. Zero does not represent the absence of effort. Zero is simply an arbitrary normalization, representing some "regular" effort level that workers would voluntarily exert even in the absence of incentives. This interpretation of the "regular" effort level as the point beyond which further work creates disutility is well known.[13]

Given that the "regular" effort level involves some effort expenditure, it is meaningful to speak of effort levels below that level. Moreover, it is natural to assume that exerting less than the "regular" effort is costly. Recalling Lazear's teaching example from the preceding footnote, most of us would not voluntarily teach five courses per semester but would continue to do some teaching even without pay. That implies that the "regular" effort level is greater than zero courses per semester and less than five courses per semester. It further implies that dropping below the regular level of courses is costly, perhaps because of boredom or feelings of guilt or diminished self worth for not furthering the educational enterprise by sharing one's knowledge.

Although values of effort below the "regular" level are valid choices, the literature leaves the cost function unspecified over the region $e_{ij}^t < 0$, and with good reason. Those effort levels are costly to the worker, and they are undesirable from the employer's standpoint, so workers will never be incentivized to choose them. Therefore, they are irrelevant to most discussions in the literature and can be safely ignored. In the present study, however, workers can enhance their promotion chances by choosing "sub-regular" effort levels (i.e.,

[12] See Prasad (2009) for another model in which workers are assumed to enter the firm with different abilities across tasks.

[13] See, for example, Holmström and Milgrom (1991). Lazear (2000) also expounds on this point on pages F614 - F615, e.g., "First, consider the statement that people are assumed to be lazy, dishonest and at odds with the goals of the managers. It is true that economists make this assumption, but they do so about behaviour on the margin. It is the marginal behaviour that is of interest to economists, and to personnel economists in particular, because the things that people want to do, do not require motivation. For example, many people take pride in their work and are willing to perform many tasks even in the absence of compensation. Most of us would continue to do some of our research and teaching even without pay. Few of us would choose to teach five courses per semester voluntarily."

strategically shirking), and that benefit must be weighed against the aforementioned costs.[14] Thus, workers are assumed to strategically shirk on a given task by purposely lowering effort on that task below the "regular" level that is normalized to zero, following the literature.[15]

Definition 1 *Worker i strategically shirks on task j in period t by choosing $e_{ij}^{t} < 0$.*

The assumption $\kappa_+ \neq \kappa_-$ allows the effort cost function to be asymmetric, though it need not be. Allowing for asymmetry is natural, because the effort costs that are relevant below the "regular" effort level (e.g., those concerning boredom, guilt, and fear of getting caught) might differ from the costs associated with productive work. To return to Lazear's teaching example, if the "regular" effort level involves teaching one course per semester, teaching zero courses or two courses would both be costly, but two courses might well be more costly than zero. The preceding cost function assumes that the "regular" effort level is unique, which is rather uncontroversial given that workers' utility functions are frequently assumed to have unique maxima. For example, an inverted-U-shaped utility is described in List et al. (2014), with a maximum at some "regular" effort level (see footnote 21, page 8), which is analogous to the present U-shaped effort cost function with a minimum at the "regular" level. Moreover, the present cost function allows for κ_- to be a considerably smaller positive number than κ_+, which allows the function to look rather flat over a range of effort values below zero.

Let $q > 0$ be a parameter that denotes the productivity of the workers' effort and that captures the relative importance of effort (as opposed to ability) in determining a worker's contribution to firm value. When employed by firm h in job k in period t, worker i's output is

$$y_{hikt} = \left(1 + s_{hi}^{t}\right)\left(c_k + d_{k\alpha} a_{i\alpha} + q e_{i\alpha}^{t} + d_{k\beta} a_{i\beta} + q e_{i\beta}^{t}\right),$$

where $d_{k\alpha} > 0$ measures the importance of ability on task α, and $d_{k\beta} > 0$ measures the importance of ability on task β in job k. Define $y_{hikt}^{\alpha} := (1 + s_{hi}^{t})(d_{k\alpha} a_{i\alpha} + q e_{i\alpha}^{t})$ and $y_{hikt}^{\beta} := (1 + s_{hi}^{t})\left(d_{k\beta} a_{i\beta} + q e_{i\beta}^{t}\right)$ as worker i's task-specific performances.[16] Besides $c_k \geq 0$, it is assumed that $d_{1\alpha} < d_{2\alpha}$ and $d_{1\beta} > d_{2\beta}$, so that task α receives more emphasis in job 2, while the opposite is true for task β. For example, task α might represent leadership activities, whereas task β might represent "actual work", i.e., the primary task

[14] The simplifying assumption that effort costs are separable by task, rather than depending on some aggregate effort level, is not critical for the qualitative results. If effort costs were dependent on the aggregate level of effort, strategic shirking would intensify. The reason is that in addition to increasing the probability of being assigned to the better-paying job, strategic shirking on one task would reduce aggregate effort and, therefore, the cost of exerting effort on the other task. Only the first of these two benefits to the worker of strategic shirking are present in the current model with a separable cost function.

A further implication of allowing effort on one task to affect the marginal cost of effort on the other task is that corner solutions may emerge. With the present cost function, the marginal cost of effort on each task is zero when that task's effort is zero. Hence, whenever effort affects the expected period-2 wage, workers exert non-zero effort. With the alternative cost function discussed in the preceding paragraph, however, if effort on task α is positive, the marginal cost of exerting effort on task β can be strictly positive even if effort on task β equals zero. If the latter cost is sufficiently high, workers may refrain from exerting effort on task β even though such effort affects the expected period-2 wage.

[15] Strategic shirking is defined as a particularly extreme form of effort misallocation. This means that efforts can be inefficiently low even if workers do not strategically shirk.

[16] One could add $z\left(1 + s_{hi}^{t}\right) c_k$ to y_{hikt}^{α} and $(1 - z)\left(1 + s_{hi}^{t}\right) c_k$ to y_{hikt}^{β}, for any $z \in [0, 1]$, without changing any of the main results, so that y_{hikt} is decomposed into the sum of two task-specific performances.

or direct production.[17] Finally, $s_{hi}^t \in \{0, S\}$, where $S > 0$, is an indicator variable capturing firm-specific human capital acquired in period 1. Its realization equals zero in period 1, or in period 2 if the worker has switched firms. The variable equals S in period 2 if the worker remains with the original employer.

It is assumed that a worker must possess experience in job 1 to perform job 2, which means that both workers are assigned to job 1 in period 1. At the end of period 1, the employer observes its workers' task-specific performances, i.e., y_{hi11}^{α} and y_{hi11}^{β}, and then assigns the workers to jobs in period 2. The incumbent firm must fill both jobs in period 2, meaning that one worker must be assigned to job 1 and the other to job 2; the case of flexible managerial job slots is covered in section 5.2. The assumption of a fixed managerial slot that must be staffed is natural and likely to be the relevant case in most organizations.[18] There is also empirical evidence consistent with fixed managerial job slots in promotion hierarchies, as discussed at the start of section 2. External employers who manage to hire a worker in period 2 are assumed, for simplicity, to always assign that worker to job 2.[19]

External firms always observe the workers' job assignments at the end of period 1. They also observe the workers' performances on tasks α and β with probabilities $p_\alpha \in [0, 1]$ and $p_\beta \in [0, 1]$, where observations of performance on tasks α and β are independent of each other. If $p_\alpha = p_\beta = 0$, learning is always asymmetric, while if $p_\alpha = p_\beta = 1$, learning is always symmetric. In all other situations, i.e., $(p_\alpha, p_\beta) \notin \{(0,0), (1,1)\}$, there is a mixture of both types of learning. These assumptions allow the models of asymmetric and symmetric learning to be merged. All firms use all available information to update their ability assessments for the workers. The workers' period-1 employer does not know whether the external firms observed the workers' period-1 performances when making job assignments at the end of period 1 and when period-2 wage offers are made.

At the beginning of period 2, the external firms attempt to hire the workers by making wage offers, whereas the current employer tries to retain the workers. The external firms move first by simultaneously making wage offers. The incumbent employer observes these wage offers and then makes a counteroffer. Each worker is hired by the firm making the highest offer. Ties are broken randomly, except when the period-1 employer is among the firms offering the highest wage, in which case the worker remains with the initial employer. S is assumed to be sufficiently high so that, in equilibrium, the external firms are never successful at hiring workers away from the period-1 employer (as seen in the forthcoming Lemma 1). As in DeVaro and Waldman (2012), however, there is a probability, γ, that the incumbent firm mistakenly fails to make a counteroffer; γ is independent of the workers' abilities and is near zero. This assumption, which is similar to Trembling-Hand Perfection (Selten 1975), simplifies the model by shifting information sets on the equilibrium path, which would otherwise be off the equilibrium path.[20]

[17] The assumption $d_{1\beta} > d_{2\beta}$ is important for the result that strategic shirking potentially occurs (see part *b* of Proposition 3). If it is instead assumed that the return to ability is higher in job 2 than in job 1 on both tasks, i.e., $d_{2\alpha} > d_{1\alpha}$ and $d_{2\beta} > d_{1\beta}$, strategic shirking would never occur (even when maintaining the assumption that $d_{2\alpha} - d_{1\alpha} > d_{2\beta} - d_{1\beta}$).

[18] Nonetheless, Waldman (2013) assumes that the managerial position can go unstaffed. Waldman and Zax (2016) also make that assumption, though they impose parametric restrictions to ensure that the firm never chooses to leave the position vacant.

[19] Most of the results do not hinge on this assumption and would be similar if external firms always assign workers that they poach to job 1, as discussed in footnote 32.

[20] The assumption that γ is independent of worker ability eliminates the strong winner's curse result that evokes Milgrom

The time structure is as follows: At the beginning of period 1, employment relationships are initiated when firms employ workers at the prevailing market wages. Workers are assigned to the low-level job, and they choose their efforts to produce period-1 outputs. At the end of period 1, their current employers observe task-specific performances and decide which worker to promote to the high-level job. The external firms observe the promotion decisions, and they observe workers' performances with certain probabilities. Then they make wage offers to the workers, to which the incumbent firms can respond. Finally, the workers choose efforts to produce period-2 outputs.

Prior to stating the model's solution, the efficient (or first-best) solution that maximizes total surplus, i.e., the difference between total output and total effort costs, is determined. This is the solution that would obtain under full information, i.e., if abilities were commonly known and effort were verifiable. The following proposition characterizes the efficient effort choices and job assignments.

Proposition 1 *a) The efficient promotion rule is based on comparative advantage. In particular, worker* 1 *(worker* 2*) is promoted to job* 2 *if and only if*

$$(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq (\leq)\, 0.$$

Because task α receives more emphasis in job 2 *than in job* 1*, the firm should promote worker* 1 *if and only if that worker has a comparative advantage on task α.*
b) Efficient efforts are $e^{fb} = \frac{(1+s^t_{hi})q}{\kappa_+}$ on both tasks, where the superscript "fb" stands either for "first best" or for "full-information benchmark". The efficient effort level equates marginal output and marginal cost.

4 Model solution

The period-1 employer observes workers' period-1 performances on both tasks (for ease of notation the firm-specific subscript is omitted, so y^j_{i11} replaces y^j_{hi11}) and then updates beliefs regarding the workers' abilities. Assume that all firms have the same belief regarding e^t_{ij} (and that all know that they share the same belief), and denote this belief by $\tilde{e}^t_{ij}$. The incumbent firm can infer worker i's ability, a_{ij}, from the output observation, y^j_{i11}, and the belief, $\tilde{e}^1_{ij}$, via the relationships[21]

$$\tilde{a}^I_{i\alpha} = \frac{y^{\alpha}_{i11} - q\tilde{e}^1_{i\alpha}}{d_{1\alpha}}$$

and Oster (1987) and that occurs in other asymmetric learning models with firm-specific human capital and counteroffers (e.g., Ghosh and Waldman 2010, DeVaro and Waldman 2012, Cassidy et al. 2016, and Waldman and Zax 2016).

[21] Since $a_{ij} \in [\underline{a}, \overline{a}]$, it would be more precise to write $\tilde{a}^I_{ij} = \max\left\{\underline{a}, \min\left\{\frac{y^j_{i11} - q\tilde{e}^1_{ij}}{d_{1j}}, \bar{a}\right\}\right\}$. This complication is caused by the limited support of the distribution and occurs only off the equilibrium path, because in equilibrium the firm correctly anticipates the workers' efforts and, thus, correctly infers the workers' ability levels. To avoid the problem, one could assume that a_{ij} is distributed on $[\underline{a}, \overline{a}]$ with probability $1-\varepsilon$ and distributed according to some other distribution with infinite support with probability ε, with $\varepsilon > 0$ and $\varepsilon \to 0$. For ease of exposition, this problem is neglected henceforth.

and

$$\tilde{a}_{i\beta}^{I} = \frac{y_{i11}^{\beta} - q\tilde{e}_{i\beta}^{1}}{d_{1\beta}},$$

with $\tilde{a}_{ij}^{I}$ denoting the firm's inference regarding a_{ij}.

External firms' expectation regarding worker i's ability on task j is denoted $\tilde{a}_{ij}^{E}$, which depends on the information about this worker's ability that the external firms observe. It is expressed as $\tilde{a}_{ij}^{E}(\sigma_\alpha, \sigma_\beta, k_i)$, where $\sigma_j \in \{j, \neg j\}$ is a binary variable that indicates whether the external firms observe the workers' performance on task j ($\sigma_j = j$) or not ($\sigma_j = \neg j$), and $k_i \in \{1, 2\}$ denotes worker i's job assignment at the end of period 1. When external firms observe workers' performances on a certain task, they can update the ability assessments for that tasks in the same way as the workers' employer. When they are unable to observe performance, they use the employer's job assignment decision to update ability assessments since that decision represents a signal about the information that the employer received.

Finally, $\tilde{a}_{ij}^{IE}(\sigma_j)$ denotes the external firms' belief regarding $\tilde{a}_{ij}^{I}$. If $\sigma_j = \neg j$, the external firms expect that y_{i11}^{j} is given by $d_{1j}a_{ij} + q\tilde{e}_{ij}^{1}$, implying $\tilde{a}_{ij}^{IE}(\neg j) = a_{ij}$. Because all firms share the same belief regarding workers' period-1 efforts, the external firms believe that the incumbent firm can correctly infer workers' ability levels. The same is true if $\sigma_j = j$. In this case, however, the external firms can even specify the exact value of $\tilde{a}_{ij}^{IE}$, since $\sigma_j = j$ means that all firms share the same information about the workers' performances on task j. Thus, $\tilde{a}_{ij}^{IE}(j) = \tilde{a}_{ij}^{I} = \frac{y_{i11}^{j} - q\tilde{e}_{ij}^{1}}{d_{1j}}$. The notation is simplified by defining $p_{\alpha\beta} := p_\alpha p_\beta$, $p_{\alpha\neg\beta} := p_\alpha(1 - p_\beta)$, $p_{\neg\alpha\beta} := (1 - p_\alpha) p_\beta$, $p_{\neg\alpha\neg\beta} := (1 - p_\alpha)(1 - p_\beta)$, $\mathbf{a} := (a_{1\alpha}, a_{1\beta}, a_{2\alpha}, a_{2\beta})$, $\Phi := (d_{2\alpha} - d_{1\alpha})(a_{2\alpha} - a_{1\alpha}) + (d_{1\beta} - d_{2\beta})(a_{1\beta} - a_{2\beta})$, $\Omega := \frac{d_{2\alpha} - d_{1\alpha}}{d_{1\beta} - d_{2\beta}}$, and by letting G denote the cdf of Φ. First-order techniques are used to determine workers' optimal efforts. It is, therefore, assumed that optimal efforts can be characterized by the first-order conditions to the workers' maximization problem and, in particular, that the objective functions are differentiable everywhere and that they are strictly concave; the latter can always be ensured by assuming a sufficiently high value for $\min\{\kappa_-, \kappa_+\}$.

The following lemma states that a sufficient amount of firm-specific human capital is needed to ensure that workers are never poached unless the incumbent firm mistakenly fails to make a counteroffer.

Lemma 1 *A threshold value $S_1 > 0$ exists such that the external firms are never successful at hiring worker i away from the incumbent firm if $S \geq S_1$, unless the incumbent firm mistakenly fails to make a counteroffer.*

The reason is that in the case of asymmetric learning, external firms condition wage offers on job assignments, because they cannot observe actual abilities. This means that in some cases, the worker's actual ability (observed by the incumbent employer) is quite low, meaning that external firms overbid, in the sense that they would prefer to lower their wage bids if they could observe actual ability. In such cases, the incumbent employer would prefer not to match offers unless firm-specific human capital is sufficiently high.[22]

[22] Such cases are more likely to occur when the ability distribution has significant dispersion, since then very low abilities (relative to external firms' wage bids) become more likely. For this reason, the threshold, S_1, is increasing in the dispersion of the ability distribution.

For tractability, it is henceforth assumed that the condition $S \geq S_1$ holds.[23] The next proposition characterizes the equilibrium.

Proposition 2 *Suppose that $S \geq S_1$. Then an equilibrium with the following properties exists:*
a) Worker Effort:
Within each period, for a given task, both workers exert the same effort level, i.e., $e^t_{1\alpha} = e^t_{2\alpha} =: e^t_\alpha$ and $e^t_{1\beta} = e^t_{2\beta} =: e^t_\beta$, for $t = 1, 2$. Moreover, in period 2, both workers exert zero effort on both tasks, i.e., $e^2_\alpha = e^2_\beta = 0$.
b) Job Assignments:
The incumbent firm promotes worker 1 to job 2 at the beginning of period 2 if and only if $(d_{2\alpha} - d_{1\alpha})\left(\tilde{a}^I_{1\alpha} - \tilde{a}^I_{2\alpha}\right) + (d_{1\beta} - d_{2\beta})\left(\tilde{a}^I_{2\beta} - \tilde{a}^I_{1\beta}\right) \geq 0$. In equilibrium, the incumbent correctly infers workers' abilities, so the inequality becomes $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0$.
c) Wages:
The period-2 wage for worker i when assigned to job k_i by the incumbent firm, and given realizations of σ_α and σ_β, is

$$w_{i2}(\sigma_\alpha, \sigma_\beta, k_i) = c_2 + d_{2\alpha}\tilde{a}^E_{i\alpha}(\sigma_\alpha, \sigma_\beta, k_i) + d_{2\beta}\tilde{a}^E_{i\beta}(\sigma_\alpha, \sigma_\beta, k_i),$$

with all $\tilde{a}^E_{ij}(\sigma_\alpha, \sigma_\beta, k_i)$ given in the appendix proof.

The intuition behind Proposition 2 is based on the relation between the external firms' belief regarding the incumbent firm's promotion rule and the incumbent firm's period-2 wage bill. Suppose that the external firms believe that the incumbent firm uses the promotion rule specified in the proposition. Then the external firms understand that the promotion decision depends only on the differences between workers' abilities on the two tasks and, importantly, that the promotion rule is fair, in the sense that either worker is promoted with probability 0.5 if both workers choose exactly the same efforts.

As a consequence of these inferences, the sum of the highest wage offers to the two workers by the external firms does not depend on which of the workers is promoted. Because the incumbent firm only has to match the highest offers to retain the workers, this implies that the promotion decision has no effect on the incumbent firm's period-2 wage bill. When making the promotion decision, therefore, the firm only considers efficiency and chooses the first-best promotion rule.[24] The result is of interest both because the assumptions that generate it are natural and empirically relevant and because the asymmetric-learning branch of the promotions literature (most of which assumes flexible slots) emphasizes inefficient job assignments.[25]

[23] Most promotion signaling models impose assumptions that ensure that there is no employee turnover (unless firms mistakenly fail to make counteroffers). See, among many other papers, Waldman (1984), Ghosh and Waldman (2010), DeVaro and Waldman (2012), and Gürtler and Gürtler (2015).

[24] The result is sensitive to the assumption that abilities are i.i.d. and that the corresponding distribution is symmetric around the mean. For example, the result might not hold in the case of heterogeneous workers with different ability distributions.

In a single-task setting, Waldman and Zax (2016) observe that the standard job-assignment inefficiency vanishes in the fixed-slot case when the managerial position is always filled. However, those authors show that when schooling is incorporated into that setting, inefficient promotions are possible of workers with higher education but lower ability than those who are not promoted.

[25] In the asymmetric-learning model of Zábojník and Bernhardt (2001), there is a single managerial slot that must be staffed,

An equilibrium is established since the external firms expect the incumbent firm to use the efficient promotion rule just described.[26] The promotion rule implies that a worker is promoted at the end of period 1 if and only if his (inferred) ability on task α is sufficiently high and his (inferred) ability on task β sufficiently low (relative to the other worker). This is intuitive. Task α is more important in job 2, whereas task β is more important in job 1. Hence, if the firm infers that a worker has high ability on task α and low ability on task β (again relative to the other worker), the firm finds it optimal to assign the worker to job 2. In the opposite case of low ability on task α and high ability on task β, the firm prefers to promote the other worker.

Workers' period-1 efforts are determined by the first-order conditions to their maximization problems. For task α, the corresponding condition is given by

$$\begin{aligned}&\kappa'\left(e_{\alpha}^{1}\right)=G'\left(0\right)\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)+p_{\alpha}\frac{d_{2\alpha}}{d_{1\alpha}}q\\&+p_{\alpha\neg\beta}G'\left(0\right)\frac{qd_{2\beta}\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\cdot\\&\left(E_{\mathbf{a}}\left[a_{i\beta}\right|a_{i\beta}\in\left(a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}\right),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\beta}\right|a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}\right)\right]\right]\right),\end{aligned}$$

with

$$W_{i2}\left(k_{i}\right):=p_{\alpha\beta}w_{i2}\left(\alpha,\beta,k_{i}\right)+p_{\alpha\neg\beta}w_{i2}\left(\alpha,\neg\beta,k_{i}\right)+p_{\neg\alpha\beta}w_{i2}\left(\neg\alpha,\beta,k_{i}\right)+p_{\neg\alpha\neg\beta}w_{i2}\left(\neg\alpha,\neg\beta,k_{i}\right)$$

and $i,l\in\{1,2\}$, $i\neq l$.

An important observation is that the expected wage spread, i.e., $E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]$, is zero in the case of pure symmetric learning, i.e., $p_{\alpha}=p_{\beta}=1$,[27] whereas it is typically non-zero when there is at least some degree of asymmetric learning, i.e., $\min\left\{p_{\alpha},p_{\beta}\right\}<1$. The reason is that when external firms perfectly observe a worker's performance on both tasks, the information about that worker's job assignment is superfluous, so external firms do not condition their wage offers on job assignments.[28] But when external employers remain in the dark about a worker's performance on at least one of the tasks, job assignments contain useful incremental information about ability, so external firms make different wage offers to promoted and non-promoted workers.

and the authors assume that the highest-ability subordinate is promoted. Therefore, job assignments are efficient. Worker investments, however, are inefficient. In that analysis the worker investments are one-dimensional human capital choices that are always inefficiently low, whereas in the present analysis they are multi-dimensional effort choices that can be either inefficiently low or high.

[26]The existence of other equilibria cannot be disconfirmed. If, for instance, the external firms anticipate that the incumbent firm uses an unfair promotion rule and, thus, favors one of the workers, the total wage bill may depend on which worker is promoted. Then, in turn, the incumbent firm may have an incentive to favor one of the workers. The multiplicity of equilibria in promotion-signaling models has been demonstrated before; see Gürtler and Gürtler (2019).

[27]The *expected* wage spread measures the difference in expected wages across jobs from the perspective of a single worker. The *actual* (i.e., observed) wage spread measures the difference between the wages of the two different workers. The actual wage spread, being dependent on the realizations of the workers' abilities, is non-zero almost surely.

[28]Formally, if $p_{\alpha}=p_{\beta}=1$, the expected wage spread becomes $E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]=E_{\mathbf{a}}\left[w_{12}\left(\alpha,\beta,2\right)\right]-E_{\mathbf{a}}\left[w_{12}\left(\alpha,\beta,1\right)\right]$. This difference can be stated as $E_{\mathbf{a}}\left[d_{2\alpha}\tilde{a}_{i\alpha}^{E}\left(\alpha,\beta,2\right)+d_{2\beta}\tilde{a}_{i\beta}^{E}\left(\alpha,\beta,2\right)-d_{2\alpha}\tilde{a}_{i\alpha}^{E}\left(\alpha,\beta,1\right)-d_{2\beta}\tilde{a}_{i\beta}^{E}\left(\alpha,\beta,1\right)\right]$ which is equal to zero since $\tilde{a}_{i\alpha}^{E}\left(\alpha,\beta,1\right)=\tilde{a}_{i\alpha}^{E}\left(\alpha,\beta,2\right)$ and $\tilde{a}_{i\beta}^{E}\left(\alpha,\beta,1\right)=\tilde{a}_{i\beta}^{E}\left(\alpha,\beta,2\right)$, as shown in the proof of Proposition 2.

Workers' incentives to exert effort on task α come from three different sources, corresponding to the three terms on the right-hand-side of the workers' first-order condition. First, as explained before, a worker is promoted at the end of period 1 if and only if, relative to the other worker, his inferred ability on task α is sufficiently high and his inferred ability on task β sufficiently low. By increasing $e^1_{i\alpha}$ the worker can signal a higher ability on task α, thereby increasing the chances of being promoted to job 2. By the same argument, the worker is more likely to be assigned to job 1 when choosing higher $e^1_{i\beta}$. Whenever expected wages differ between the two jobs (i.e., $E_{\mathbf{a}}[W_{12}(2)] \neq E_{\mathbf{a}}[W_{12}(1)]$), workers have an incentive to exert effort to affect their period-2 job assignments in order to be assigned to the better-paying job. Second, conditional on being assigned to a specific job k, a worker's effort on task j affects the worker's compensation only if the external firms are able to observe the worker's performance on that task. If task-specific performance is observable, by choosing higher effort, the worker signals higher ability on the corresponding task, rendering him more attractive to the external firms. Third, by exerting effort on one task, workers can favorably affect the ability assessment for the other task if performance on the former task is observable and performance on the latter task is unobservable. Suppose, for example, that performance on task α is observable, whereas external firms cannot observe performance on task β. By increasing effort on task α, a worker increases the external firms' ability assessment on that task. Recall that the worker is promoted to job 2 if and only if his ability on task α is high and his ability on task β low (relative to the other worker). Hence, if the external firms observe the worker's performance on task α and infer that the worker has a high ability on that task, they conclude that the worker is promoted to job 2 for a wider range of ability levels on task β. This implies that external firms downgrade the worker's ability assessment for task β less strongly when they observe the worker's promotion and when they are unable to observe the worker's performance on task β. Likewise, they upgrade the worker's ability assessment for task β to a greater degree when they observe that the worker is not promoted. The effort choice is symmetric because workers have the same effort cost functions and choose efforts before learning their (potentially different) abilities.

The first-order condition for task β is given by

$$\kappa'\left(e^1_\beta\right) = -G'(0)\,\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}(2)\right]-E_{\mathbf{a}}\left[W_{12}(1)\right]\right)+p_\beta\frac{d_{2\beta}}{d_{1\beta}}q$$
$$+p_{\neg\alpha\beta}G'(0)\,\frac{qd_{2\alpha}\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\cdot$$
$$\left(E_{\mathbf{a}}\left[a_{i\alpha}\middle|\,a_{i\alpha}\in\left(a_{l\alpha}+\frac{1}{\Omega}\left(a_{i\beta}-a_{l\beta}\right),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\alpha}\middle|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}\left(a_{i\beta}-a_{l\beta}\right)\right]\right]\right),$$

and the interpretation is analogous to that for task α.

In the next step, workers' optimal efforts are further analyzed. In particular, the workers are shown to potentially have an incentive to strategically shirk.

Proposition 3 *In the equilibrium described in Proposition 2:*

a) If $p_\alpha = p_\beta = 1$, then $e^1_\alpha > e^{fb} > e^1_\beta > 0$.

b) There exists (p_α, p_β) *with* $(p_\alpha, p_\beta) \neq (1,1)$ *such that* $e^1_\alpha < 0$ *or* $e^1_\beta < 0$*, but not both simultaneously.*

The intuition behind part *a*) of Proposition 3 is as follows. If learning is entirely symmetric, the expected wage spread from promotion is zero, so the workers have no incentive to manipulate their period-1 effort choices to influence the promotion probability. However, the external firms observe workers' task-specific performances and use these performances to update expectations about worker ability. Therefore, the workers have an incentive to provide good performance signals on both tasks and, thus, choose positive effort. The effect is such that workers exert inefficiently high effort on task α, but inefficiently low effort on task β.

To understand why efforts are inefficient when learning is symmetric, notice that, when the firms deduce workers' abilities on a task from the performance observations, they divide performances by d_{1j}. This implies that the marginal effect of effort on the ability assessment is given by $\frac{q}{d_{1j}}$. Furthermore, since workers are assigned to job 2 when being poached by an external firm, their ability assessments for task j are multiplied by d_{2j} in the period-2 wage function. This in turn means that the marginal effect of period-1 effort on task j on the period-2 wage is $q\frac{d_{2j}}{d_{1j}}$. Given that $q\frac{d_{2\alpha}}{d_{1\alpha}} > q > q\frac{d_{2\beta}}{d_{1\beta}}$, the efforts are inefficiently high on task α and inefficiently low on task β.

The intuition behind part *b*) of Proposition 3 is linked to that of Proposition 2. In Proposition 2 workers' period-2 job assignment depends on the incumbent firm's inference regarding workers' abilities. Whenever $\min\{p_\alpha, p_\beta\} < 1$, external firms account for job assignments when updating their expectation about worker ability. It follows that workers' expected wages may differ across jobs, in which case the workers wish to affect job assignments so as to be assigned to the better-paying job in period 2. Hence, workers distort their efforts so as to change the incumbent firm's inference regarding their abilities. For instance, if the expected wage is higher in job 2 than in job 1, workers have an incentive to increase effort on task α, but to lower effort on task β to change the incumbent firm's ability assessments on the two tasks and to increase the probability of being assigned to job 2. The latter effect may be so strong that optimal effort on task β becomes negative, i.e., workers strategically shirk on that task. Therefore, asymmetric learning is a necessary (but not sufficient) condition for strategic shirking to occur.

In principle, the expected wage in job 1 might exceed the expected wage in job 2, in which case strategic shirking could occur on task α. The reason is that promotion to job 2 leads to an upgrade in the external firms' ability assessment on task α and a simultaneous downgrade in the corresponding assessment for task β. If the latter effect is strong enough, the offered wage may decline.

In practice, however, a promotion is typically expected to lead to an increase in the wage. By imposing additional restrictions on the model parameters, the expected period-2 wage can always be assured to be higher in job 2 than in job 1. For instance, assume that $p_{\neg\alpha\beta}$ is high, i.e., it is difficult for the external firms to observe performance on task α, whereas they are likely to receive information regarding the performance on task β. Then a situation is likely in which the exteral firms do not observe performance on task α, so that a promotion leads to an upgrade in the ability assessment for task α, while they observe performance on task β, so that the promotion does not lead to a corresponding downgrade in the ability assessment for task β. When $p_{-\alpha\beta}$ is high enough, this effect dominates all other effects on the period-2 wage, implying that the expected period-2 wage is higher in job 2 than in job 1.[29]

Proposition 3 shows that the parameters p_α and p_β affect the optimal effort levels. The next proposition sharpens this observation and highlights the welfare implications of changes in p_α or p_β.

Proposition 4 *In the equilibrium described in Proposition 2:*

a) There are necessary and sufficient conditions, stated in the appendix proof, under which period-1 worker efforts on a given task are increasing in the probability that competing firms in the market can observe worker performance on that task, i.e., $\frac{\partial e_\alpha^1}{\partial p_\alpha} > 0$ *and/or* $\frac{\partial e_\beta^1}{\partial p_\beta} > 0$.

b) Period-1 worker efforts on a given task are non-decreasing in the probability that competing firms in the market can observe worker performance on the other task, i.e., $\frac{\partial e_\alpha^1}{\partial p_\beta} \geq 0$ *and* $\frac{\partial e_\beta^1}{\partial p_\alpha} \geq 0$*. Moreover,* $\frac{\partial e_\alpha^1}{\partial p_\beta} > 0$ *when* $p_\alpha < 1$*, and* $\frac{\partial e_\beta^1}{\partial p_\alpha} > 0$ *when* $p_\beta < 1$*.*

c) It is possible that an increase in p_j *reduces welfare by moving effort farther away from the efficient level. In particular, if an additional condition, stated in the appendix proof, is met, welfare is not maximized at* $(p_\alpha, p_\beta) = (1, 1)$*.*

To understand the intuition for Proposition 4, consider, first, a marginal increase in p_α, which affects the incentive to exert effort in three ways. First, the external firms are more likely to observe the workers' performance on task α, giving the workers a stronger incentive to exert effort on that task. Second, the expected wage spread, $E_{\mathbf{a}}[W_{12}(2)] - E_{\mathbf{a}}[W_{12}(1)]$, decreases. The reason is that a wage spread obtains only because external firms are sometimes unable to observe workers' performances, thus conditioning their wage offers on the incumbent firm's promotion decision. Because a worker with a high ability on task α is more likely to be promoted than one with low ability, external firms' wage offers are, *ceteris paribus*, higher for promoted workers than for non-promoted ones when performance on task α is unobservable. An increase in p_α diminishes this effect, and the expected wage in job 2 becomes relatively lower compared to that in job 1. Workers, therefore, have a stronger incentive to be assigned to job 1, implying a weaker incentive to exert effort on task α and a stronger incentive to exert effort on task β. Third, as explained before, by exerting effort on one task, workers can favorably affect the ability assessment for the other task if performance on the former task is observable and performance on the latter task is unobservable. When p_α increases, a situation becomes more likely in which performance on task α is observable, whereas performance on task β cannot be observed by the external firms. Likewise, a situation in which performance on task β is observable, whereas performance on task α cannot be observed by the external firms, is less likely to occur. This increases the incentive to exert effort on task α, while decreasing the incentive to exert effort on task β. In sum, an increase in p_α has three different effects on the incentive to exert effort on task α, meaning that this effort

[29] One feature of the current model (and the one by Waldman 1984) is that wages can decrease over time, meaning that the period-1 wage can be higher than the (expected) period-2 wage. This result conflicts with real-world compensation schemes which tend to be increasing over time. One way to reconcile these results with the empirical evidence is to assume that workers acquire general human capital (in addition to firm-specific human capital) in period 1, which makes them more productive at other firms in period 2. Due to general human capital acquisition, period-2 wages would increase over time.

can either increase or decrease. The condition ensuring $\frac{\partial e_{\alpha}^{1}}{\partial p_{\alpha}} > 0$ is given by

$$\frac{d_{2\alpha}}{d_{1\alpha}}q - \frac{d_{2\alpha}}{d_{1\alpha}}qG'(0)\,(d_{2\alpha} - d_{1\alpha})\,E_{\mathbf{a}}\left[a_{1\alpha} - a_{2\alpha}|\, a_{1\alpha} - a_{2\alpha} \geq \frac{1}{\Omega}\,(a_{1\beta} - a_{2\beta})\right] +$$
$$q\,(1 - p_{\beta})\,G'(0)\,\frac{d_{2\beta}}{d_{1\alpha}}\,(d_{2\alpha} - d_{1\alpha})\,E_{\mathbf{a}}\,[a_{1\beta} - a_{2\beta}|\, a_{1\beta} - a_{2\beta} \geq \Omega\,(a_{1\alpha} - a_{2\alpha})] > 0,$$

with the first term representing the stronger incentive to exert effort because of the greater probability that external firms observe the workers' performance on task α, the second term the weaker incentive to exert effort due to the change in the expected wage spread, and the third term the stronger incentive to exert effort to favorably affect the ability assessment on task β if performance on task α is observable and performance on task β is unobservable. Finally, while an increase in p_{α} has two opposing effects on the incentive to exert effort on task β, one can show that the positive effect dominates, so that effort on task β always increases.[30]

The effects of a marginal increase in p_{β} on the optimal efforts are similar. First, workers have a stronger incentive to exert effort on task β because the external firms are more likely to observe their performance on that task. Second, the expected wage in job 1 becomes relatively lower compared to that in job 2, so that the expected wage spread increases.[31] Workers, therefore, have a stronger incentive to be assigned to job 2, implying a stronger incentive to exert effort on task α and a weaker incentive to exert effort on task β. Third, a situation becomes more likely in which performance on task β is observable, whereas performance on task α cannot be observed by the external firms. Likewise, a situation in which performance on task α is observable, whereas performance on task β cannot be observed by the external firms, is less likely to occur. This increases the incentive to exert effort on task β, while decreasing the incentive to exert effort on task α, as previously explained. Considering all of these effects, the effort on task α always increases, whereas the effort on task β may increase or decrease.

The result that an increase in p_{α} or p_{β} can be welfare-reducing can be explained as follows. Effort on task α is inefficiently high in the case of pure symmetric learning (i.e., $p_{\alpha} = p_{\beta} = 1$), whereas effort on task β is inefficiently low, as explained before. A decrease in p_{α} can now lead to a lower incentive to exert effort on task α. Therefore, effort may become closer to the efficient level so that welfare would increase. Similarly, the incentive to exert effort on task β may increase as p_{β} becomes lower, which would again lead to higher welfare.[32]

[30] These results differ substantially from those in Holmström and Milgrom (1991). In that model, performance measures for different tasks are contractible so that they can be used in formal incentive contracts. If a performance measure on one task becomes more precise, the firm puts greater emphasis on this measure in the incentive contract. Hence, the worker finds it optimal to exert higher effort on that task and, if task-specific efforts are substitutes, reduces the effort on the other task.

[31] Again, the expected wage spread can be negative, in which case an increase in the expected wage spread means that the absolute value of the expected wage spread shrinks.

[32] The preceding analysis assumes that workers who switch firms are assigned to job 2 in the new firm. Under the alternative assumption that such workers are assigned to job 1, efficient efforts would result in the case of pure symmetric learning. The reason is that ability assessments for the two different tasks would be multiplied by $d_{1\alpha}$ and $d_{1\beta}$ in the period-2 wage function, implying that the marginal effect of period-1 effort on the period-2 wage would be q. Still, one can show that an increase in one of the p_j's could be welfare-reducing, at least locally. A formal proof of this statement is available upon request.

4.1 Initial matching of workers to firms

This section addresses the period-1 matching of workers to firms and determines the period-1 wage, w_1. Let n_h denote the number of workers that firm h hires in period 1. It is well known that firms' equilibrium labor demand (for given wage w_1) equals the labor demand that maximizes the firms' aggregate profits (see, e.g., Proposition 5.E.1 in Mas-Colell et al. 1995). Labor supply is fully inelastic in the present model, implying that firms' total wage costs are fixed (at $\sum_{h=1}^{N} n_h w_1 = nw_1$). It follows that equilibrium labor demands maximize the firms' total expected output (net of period-2 wage costs). Expected output per worker is assumed to be highest if a firm hires two workers. Since the number of workers is fixed, total expected output is highest if $n/2$ of the firms hire two workers each, and such allocation of workers to firms will result in equilibrium.

The period-1 wage is determined such that firms' expected profit over both periods is zero. If it were higher, the firms hiring workers would make a loss, and they would gain from deviating to a lower wage offer. If it were lower, profits would be positive, and the firms that failed to hire workers would deviate to a (slightly) higher wage. The period-1 wage is thus given by

$$\begin{aligned} w_1 &= c_1 + (d_{1\alpha} + d_{1\beta}) E[a] + q\left(e_\alpha^1 + e_\beta^1\right) - 0.5\left(E_{\mathbf{a}}[W_{12}(2)] + E_{\mathbf{a}}[W_{12}(1)]\right) \\ &+ 0.5(1+S)\left(c_2 + E_{\mathbf{a}}\left[d_{2\alpha}a_{1\alpha} + d_{2\beta}a_{1\beta} \middle| (d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0\right]\right) \\ &+ 0.5(1+S)\left(c_1 + E_{\mathbf{a}}\left[d_{1\alpha}a_{1\alpha} + d_{1\beta}a_{1\beta} \middle| (d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \leq 0\right]\right). \end{aligned}$$

If a firm hires two workers in period 1, it earns positive period-2 profit due to human-capital accumulation. Therefore, firms are willing to incur a loss in period 1, offering workers wages exceeding their expected output.

The assumption that output (net of period-2 wages) is maximized if firms hire exactly two workers is made for simplicity and should not be crucial for the results. To see this, suppose that firms find it optimal to hire more than two workers and that they continue to promote only one of these workers. The promotion decision will still be based on comparative advantage. Accordingly, workers face very similar trade-offs as in the two-worker case and, importantly, strategic shirking remains an attractive option in some situations.

The assumption of two periods further simplifies the analysis, but strategic shirking arises even in settings with more than two periods. As an illustrative example, suppose that there are three periods and that two workers are hired into a low-level job in period 1. One of the workers is promoted to a middle-level job in period 2, and this worker has the opportunity to further advance to a high-level job in period 3. If the two tasks differ in importance across jobs and, say, task α becomes ever more important as a worker moves up the corporate ladder, workers might again have an incentive to strategically shirk on task β to gain an advantage in the period-1 competition.

5 Extensions

In the model, the information that competing employers receive about performance is task-specific. But in some settings, external firms might observe only a single, aggregate performance measure. The results would remain largely unchanged in such a setting, though two things should change. First, the third source of worker incentives discussed earlier (i.e., by exerting effort on one task, workers can favorably affect the ability assessment for the other task if performance on the former task is observable and performance on the latter task is unobservable) should disappear, because this effect requires that performance on exactly one of the tasks is observable. Second, observation of an aggregate performance measure would not allow external firms to perfectly infer workers' abilities, because the aggregate measure would be an amalgamation of the performances on both tasks.

Although the main model incorporates task-specific efforts and abilities, some models in the literature omit either or both of these variables, as indicated in the typology discussed in the introduction. Such models can be understood as limiting cases of the general model. For example, if q approaches zero, the incentive to exert effort vanishes. Similarly, if $\overline{a}$ approaches $\underline{a}$, workers would have essentially the same ability, so that ability would effectively vanish from the model. Analyses of these special cases are omitted but are available upon request. Three extensions are explored next. The first is a single-task model, as in the bulk of the theoretical promotions literature. The second considers flexible managerial job slots, and the third is a classic tournament. The section concludes with a discussion of the potential solutions to the strategic-shirking problem.

5.1 A single-task model

Single-task models dominate both branches of the promotions literature discussed in the introduction, so they offer a natural benchmark case against which to compare the main analysis. As was shown in the multi-task case, when there are two tasks, strategic shirking on one task can be counterbalanced by working extra hard on the other task, so that it is still possible to impress the boss while manipulating future job assignments. But if there is only one task, workers cannot shirk without fearing a bad outcome, so workers never strategically shirk on primary tasks. The reason for this result, under both assumptions about learning, is as follows.

If learning is symmetric, the expected wage spread between the two jobs is zero, so workers have no incentive to strategically shirk on primary tasks. The expected wage spread is zero because all employers observe the workers' performances on both tasks, so job assignments provide no further useful information; competing employers, therefore, do not condition their wage offers on observed job assignments. In contrast, if learning is asymmetric, job assignments provide incremental information about ability, so competing employers condition their wage offers on job assignments, and the expected wage spread is nonzero. But strategic shirking on primary tasks is unappealing, because it would increase a worker's probability of getting assigned to the lower-paying job 1.

When the single-task learning environment becomes increasingly symmetric on primary tasks, worker effort could either increase or decrease, and welfare could again decrease. Three effects were mentioned in the multi-task case, and the first two of those effects also apply to the single-task case.

To formalize the preceding results, suppose without loss of generality, that the workers only perform task α. This is modeled by assuming that task β does not exist, so that all variables relating to that task are set equal to zero. Suppose also that $S \geq S_1$. The following proposition characterizes the equilibrium, where $\hat{\Phi} := (d_{2\alpha} - d_{1\alpha})(a_{2\alpha} - a_{1\alpha})$, $\hat{\mathbf{a}} := (a_{1\alpha}, a_{2\alpha})$, and $\hat{G}$ denotes the cdf of $\hat{\Phi}$.

Proposition 5 *If workers perform only task α, and $S \geq S_1$, then an equilibrium with the following properties exists:*

a) Worker Effort:

Within each period, both workers exert the same effort level, i.e., $e^t_{1\alpha} = e^t_{2\alpha} =: e^t_\alpha$, for $t = 1, 2$. In period 1, these efforts are determined by

$$\kappa'\left(e^1_\alpha\right) = \hat{G}'(0)\,\frac{q\,(d_{2\alpha} - d_{1\alpha})}{d_{1\alpha}}\left(E_{\hat{\mathbf{a}}}\left[W_{12}(2)\right] - E_{\hat{\mathbf{a}}}\left[W_{12}(1)\right]\right) + p_\alpha \frac{d_{2\alpha}}{d_{1\alpha}} q, \text{ with}$$

$$W_{i2}(k_i) := p_\alpha w_{i2}(\alpha, k_i) + (1 - p_\alpha)\, w_{i2}(\neg\alpha, k_i)\,.$$

In period 2, both workers exert zero effort, i.e., $e^2_\alpha = 0$.

b) Job Assignments:

The incumbent firm promotes worker 1 to job 2 at the beginning of period 2 if and only if $\tilde{a}^I_{1\alpha} \geq \tilde{a}^I_{2\alpha}$. In equilibrium, the incumbent correctly infers workers' abilities, so the inequality becomes $a_{1\alpha} \geq a_{2\alpha}$.

c) Wages:

The period-2 wage for worker i when assigned to job k_i by the incumbent firm, and given realization of σ_α, is

$$w_{i2}(\sigma_\alpha, k_i) = c_2 + d_{2\alpha}\tilde{a}^E_{i\alpha}(\sigma_\alpha, k_i)\,,$$

with all $\tilde{a}^E_{i\alpha}(\sigma_\alpha, k_i)$ stated in the appendix proof.

The period-1 wage is determined such that firms earn zero expected profit. It is also stated in the appendix proof.

Proposition 5 has the following corollary:

Corollary 1 *In the equilibrium from Proposition 5 workers' efforts always satisfy $e^1_\alpha > 0$.*

In contrast to the main model, in the single-task model a promotion always entails a wage increase. The reason is that, except in the case of pure symmetric learning, a promotion always induces the external firms to upgrade the ability assessment for the promoted worker on the single task, thus implying a higher wage offer. In the case of pure symmetric learning, effort increases the performance measure that all firms observe, which implies a higher wage offer. A consequence is that workers' effort is always positive, regardless of the value of p_α. Strategic shirking is never an attractive option for a worker, because it increases the probability that the worker is assigned to the low-paying job 1 if the external firms do not observe the worker's performance, while decreasing the worker's ability assessment, and thus the wage, if the external firms observe the worker's performance. Thus, multiple tasks are necessary for strategic shirking to occur.

Next, consider whether effort increases or decreases when learning becomes more symmetric. Again it is possible that an increase in p_α reduces welfare.

Proposition 6 *In the equilibrium from Proposition 5:*
a) $\frac{\partial e_\alpha^1}{\partial p_\alpha} > 0 \Leftrightarrow 1 > \hat{G}'(0)(d_{2\alpha} - d_{1\alpha}) E_{\hat{\mathbf{a}}}\left[a_{1\alpha} - a_{2\alpha} | a_{1\alpha} - a_{2\alpha} \geq 0\right].$
b) If $1 > \hat{G}'(0)(d_{2\alpha} - d_{1\alpha}) E_{\hat{\mathbf{a}}}\left[a_{1\alpha} - a_{2\alpha} | a_{1\alpha} - a_{2\alpha} \geq 0\right]$ *and* p_α *increases, it is possible that* e_α^1 *moves farther away from the efficient level. In particular, welfare is not maximized at* $p_\alpha = 1$.

An inefficiently high effort on task α is chosen if $p_\alpha = 1$. Hence, if a reduction in p_α leads to a lower effort, the effort gets closer to the efficient level and welfare increases. Effort is inefficient in the asymmetric learning model for the following reason. Workers exert effort because doing so increases the probability of promotion and an accompanying wage increase. In Lazear and Rosen (1981), if workers are risk neutral, the firm sets the wage spread to induce the efficient effort choice. In the present model, however, no firm is able to optimally choose a wage spread. Instead, wages and the wage spread are determined by labor market competition for the workers' services. Therefore, the efficient wage spread could result by coincidence, but there is no reason to expect this to occur. Rather, it is likely that an inefficiently low or high wage spread results.

The single-task version of the model is similar to the model in section 4 of Ghosh and Waldman (2010) when firms choose standard (as opposed to up-or-out) promotion practices. Some differences are that the present model has two competing workers and one managerial job that must be staffed, whereas their model has a single worker who may or may not get promoted; a winner's curse result determines wages in their model but not in the present one; ability is binary in their model and continuous in the present one; worker output has a stochastic component in their model but not in the present one; learning is asymmetric in their model, whereas in the present model a range of possibilities for learning (including the extreme cases of symmetric and asymmetric learning) are considered. Moreover, the focus of that analysis is on the firm's choice between standard versus up-or-out promotion rules, whereas the focus of the present analysis is on the efficiency of effort allocation.

5.2 Flexible managerial job slots

The main model assumes a job slot constraint at the top of the hierarchy, so that there is only one managerial position. This is the most relevant case in most organizations, and as discussed at the start of section 2, it is consistent with empirical evidence that relative worker performance determines promotions, which is an implication of managerial slot constraints. Nonetheless, much of the promotions literature that is based on job-assignment models assumes flexible managerial job slots.[33] This section summarizes how the present analysis changes in that environment. The formal results are available upon request.

The only difference between a model with flexible job slots and the main model with fixed slots is that under flexible slots the promotion (or job assignment) rule can be inefficient, so the celebrated result of Waldman (1984) in a single-task setting is preserved in a multi-task one. In contrast, under fixed slots when the managerial job must be staffed, the promotion rule is efficient, which is important given the empirical relevance of the fixed-slot case. All other results are qualitatively the same between the two cases. An implication of potential inefficiency in job assignments is that if the learning environment shifts in the direction of symmetric learning, the welfare analysis involves an increase in efficiency in job assignments along with the ambiguous change in efficiency in effort allocation that was present in the main model with fixed managerial job slots.

5.3 Classic tournaments

Most models of classic tournaments assume a single firm that employs two competing workers, and the single-firm assumption is also adopted in this subsection. Before the results are presented, a brief explanation is provided for how the single-firm assumption would change the results from the main model in which commitment to future wages is not possible. The firm would always offer both workers a period-2 wage equal to their reservation value, implying a wage spread of zero. Workers anticipate that their period-1 performance has no effect on their period-2 compensation and, therefore, choose zero effort in both periods. While workers have no incentive to choose effort higher than the regular level, strategic shirking is not an attractive option either. Furthermore, the firm chooses the same promotion rule as in the main model. Because period-2 wage costs do not depend on which worker is promoted, the firm assigns workers to the two jobs so as to maximize their total output. All of these results would also obtain in the case of two firms. A proof is available upon request.

The analysis of the classic tournament is very similar to the one in DeVaro and Gürtler (2016a). The firm offers worker i a contract which pays the worker a period-2 wage of $w_2(k_i)$, which may depend on the worker's period-2 job assignment. The firm uses the same promotion rule as in the main model, and workers' period-1 efforts are determined by the optimality conditions

$$G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}(w_2(2)-w_2(1))=\kappa'\left(e_\alpha^1\right)$$

and

$$-G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}(w_2(2)-w_2(1))=\kappa'\left(e_\beta^1\right).$$

[33] See, for example, Waldman (1984), Prendergast (1993), Bernhardt (1995), Gibbons and Waldman (1999, 2006), Ghosh and Waldman (2010), DeVaro and Waldman (2012), Cassidy et al. (2016), and DeVaro et al. (2018).

The conditions are similar to those from the main model, with the two exceptions that the expected wages in the different job levels ($E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]$ and $E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]$) are replaced by the actual wages, $w_2\left(1\right)$ and $w_2\left(2\right)$, that the firm sets in advance, and that the remaining terms disappear from the conditions since there are no external firms. If the firm sets $w_2\left(1\right) = w_2\left(2\right)$, workers choose zero effort on both tasks. In contrast, if the firm sets $w_2\left(1\right) \neq w_2\left(2\right)$, workers exert positive effort on one of the tasks, while strategically shirking on the other. The wages, $w_2\left(1\right)$ and $w_2\left(2\right)$, are chosen by the firm to resolve this trade-off optimally. In general, the firm does not implement efficient effort levels.

5.4 Solutions to the strategic-shirking problem

The organizational costs of strategic shirking create a motivation for firms to devise ways to combat the problem. A possible solution is for the employer to commit to publicizing workers' performances, because strategic shirking occurs only under asymmetric learning. This solution is particularly interesting, because it runs counter to the natural intuition that employers should try to preserve, and capitalize on, their private information about their workers' talents. Examples abound in academia, law, consulting, and elsewhere, in which employers showcase their workers' talents on the company website. In most cases such public displays are probably intended for customers rather than for competitors. But they can be seen by competitors nonetheless, which implies that a byproduct of such advertising is a reduction in incentives to strategically shirk.

The preceding argument is easily formalized. Suppose that in the main model external firms observe workers' performances on task α with exogenously given probability p_α, whereas the incumbent firm partly controls the flow of information regarding task β. More specifically, suppose that external firms observe performance on task β with exogenously given probability $p_\beta < 1$ unless the incumbent firm actively discloses the corresponding performance information, in which case $p_\beta = 1$. The incumbent firm is assumed to disclose performance information on task β if and only if this leads to an increase in total surplus. Increasing surplus is optimal for the firm because the firm is always able to allocate the additional surplus between itself and the two workers in a way that makes all three parties better off. Since job assignments are efficient regardless of the exact value of p_β and period-2 efforts are always zero, the firm discloses performance information on task β if and only if this increases the value of

$$q\left(e_\alpha^1 + e_\beta^1\right) - \kappa\left(e_\alpha^1\right) - \kappa\left(e_\beta^1\right).$$

Suppose that

$$\begin{aligned} & \frac{d_{2\beta}}{d_{1\beta}} + \left(1 - p_\alpha\right) G'\left(0\right) \frac{d_{2\alpha}\left(d_{1\beta} - d_{2\beta}\right)}{d_{1\beta}} E_{\mathbf{a}}\left[a_{1\alpha} - a_{2\alpha} \middle| \, a_{1\alpha} - a_{2\alpha} \geq \frac{1}{\Omega}\left(a_{1\beta} - a_{2\beta}\right)\right] \\ > \quad & G'\left(0\right)\left(d_{1\beta} - d_{2\beta}\right) \frac{d_{2\beta}}{d_{1\beta}} E_{\mathbf{a}}\left[a_{1\beta} - a_{2\beta} \middle| \, a_{1\beta} - a_{2\beta} \geq \Omega\left(a_{1\alpha} - a_{2\alpha}\right)\right]. \end{aligned}$$

From the proof of Proposition 4 it is then clear that e_β^1 gets closer to the efficient level as p_β increases (which implies that $qe_\beta^1 - \kappa\left(e_\beta^1\right)$ is maximized at $p_\beta = 1$). In addition, using the results from Proposition 2 one can show that, for $p_\beta = 1$, effort on task α is characterized by

$$\kappa'\left(e_\alpha^1\right) = G'(0)\frac{d_{2\alpha}}{d_{1\alpha}}q\left(d_{2\alpha} - d_{1\alpha}\right)\left(1 - p_\alpha\right)E_{\mathbf{a}}\left[a_{1\alpha} - a_{2\alpha}\middle| a_{1\alpha} - a_{2\alpha} \geq \frac{1}{\Omega}\left(a_{1\beta} - a_{2\beta}\right)\right] + p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q.$$

So if

$$G'(0)\frac{d_{2\alpha}}{d_{1\alpha}}q\left(d_{2\alpha} - d_{1\alpha}\right)\left(1 - p_\alpha\right)E_{\mathbf{a}}\left[a_{1\alpha} - a_{2\alpha}\middle| a_{1\alpha} - a_{2\alpha} \geq \frac{1}{\Omega}\left(a_{1\beta} - a_{2\beta}\right)\right] + p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q \leq q,$$

then even for $p_\beta = 1$, effort on task α is not inefficiently high which, again, using the results from Proposition 4, shows that

$$qe_\alpha^1 - \kappa\left(e_\alpha^1\right)$$

is also maximized at $p_\beta = 1$. In this case, the firm would clearly benefit from committing to reveal information about its workers' performances on task β. If the above inequality is not met, a higher p_β may lead to a decrease in $qe_\alpha^1 - \kappa\left(e_\alpha^1\right)$. The firm would then have to trade off the increase in surplus on task β and the decrease in surplus on task α when deciding whether or not to commit to $p_\beta = 1$.

Job design offers another solution to address the strategic-shirking problem. The analysis treats the grouping of tasks into jobs, within and across hierarchical levels, as predetermined, but employers can design jobs with an eye towards lessening the strategic-shirking problem. Given that strategic shirking does not occur in single-task settings, assigning one task to each worker would be efficient, when technologically feasible.[34][35] Yet another solution involves the employer committing to a more balanced promotion rule (i.e., one that places greater weight on the task that is emphasized in the lower-level job). Such commitment is disallowed in the present analysis but may be possible in some settings, as discussed in DeVaro and Gürtler (2016a); moreover, Benson et al. (2019) provide empirical evidence from the sales industry suggesting that firms can commit to promotion rules that limit the inefficiencies associated with strategic shirking.

6 Empirical implications

The model integrates both pillars of the theoretical promotions literature, generalized to a multi-task environment. Either of those pillars, taken alone, has empirical implications that extend to the present model. Rather than recapitulating these, the present section focuses on new ones.

[34] Similarly, Holmström and Milgrom (1991) argue, in the case of jobs in which one task is easy to measure and the other is difficult to measure, that the tasks should be separated into distinct jobs, with strong incentives attached to the task that can be measured easily.

[35] Large firms may be more technologically able to specialize workers, whereas in small firms, workers are more often required to assume more job tasks (e.g., there may be enough human resources work in a large firm for someone to fully specialize in that, whereas in a small firm the human resources person must also handle other tasks, thereby increasing the likelihood of strategic shirking). An empirical relationship between large firm size and narrow job design is predicted.

The first set of empirical implications does not require multiple tasks, endogenous effort choices, or even an integrative model, but merely the observation that when there are fixed managerial slots that must be staffed, no inefficient job assignments occur under asymmetric learning. A growing empirical literature that relies on asymmetric learning models with flexible job slots has been based on how the degree of inefficiency in job assignments varies with observable characteristics such as educational attainment, gender, race, and age.[36] All of those empirical tests have the further implication that in data samples where flexible managerial slots can reasonably be assumed, the inefficiency should vary with observable characteristics in the manner posited in those studies, whereas in samples where a fixed-slot assumption is more reasonable, that variation should be muted.[37]

6.1 Overlapping wage distributions across adjacent job levels

The model predicts overlapping wage distributions across adjacent job levels, which is an important stylized fact both in personnel data (e.g., Baker et al. 1994) and in matched worker-firm panel data (e.g., Kauhanen and Napari 2012) that has been a challenge to capture convincingly in some influential theoretical models. For example, in their footnote 13, Gibbons and Waldman (1999) note that their model cannot produce the empirical prediction.

Gibbons and Waldman (2006) incorporate schooling into their 1999 model and show that it can explain wage overlap across adjacent job levels.[38] The schooling extension generates wage overlap across levels only because workers of different schooling levels are mixed together within the job hierarchy. Within an education level, the model generates no wage overlap across job levels. In earlier unpublished analysis using the personnel data from Baker et al. (1994), DeVaro found, in collaboration with Hugh Cassidy, that variation in education levels is not what drives wage overlap across job levels. The results of that analysis are summarized here; further details are available upon request.

The analysis used the same four education groups that were used in DeVaro and Waldman (2012). These were derived using workers' years of education and were based on the typical years required to complete a given degree.[39] The bulk of the personnel data are contained in the lowest three job levels of the white-collar

[36]See, for example, Okamura (2011), DeVaro and Waldman (2012), Bognanno and Melero (2016), Cassidy et al. (2016), and DeVaro et al. (2018).

[37]Waldman and Zax (2016) note that there are two types of inefficiency in job assignments that potentially occur under asymmetric learning. One concerns the number of workers who get promoted, and another concerns the type of worker (e.g., highly educated or less educated) that gets promoted. The former inefficiency vanishes when there is a fixed managerial job slot that must be staffed, and only the latter remains, so overall inefficiency in job assignments should be muted.

[38]The following quote from their 1999 paper discusses an extension of their 1999 model that would generate wage overlap across job levels, though in their 2006 paper they argue in footnote 11 that their schooling-based argument from 2006 is superior: "BGH's empirical analysis and our theoretical analysis treat the firm as if there were a single job ladder. But for most large firms there are multiple job ladders, such as those associated with different occupations. Given this, suppose we extended our model so that the firm has two job ladders but promotions occur only within ladders, not across. Suppose further that promotions occur only from the top of the wage distribution at one level of a ladder to the bottom of the wage distribution at the next level of that ladder. If the wage distributions for one ladder are shifted relative to those of another (say, because of compensating differentials or other occupational wage differences), then an analysis of the aggregate job ladder would yield overlapping wage distributions." (p. 1341)

[39]High school graduates were defined to have 12, 13, or 14 years of education; bachelor's degree holders were defined to have 16 years; master's degree holders were defined to have 18 years; PhD holders were defined to have 21 or more years. See DeVaro and Waldman (2012) for further discussion.

portion of the hierarchy. The following approach was used to approximate the degree of wage overlap across job levels.

First, kernel densities for the wage distributions in job level 1 (the lowest level) and job level 2 (the next level up) were estimated, and overlayed.

Second, the supports of those distributions were partitioned into discrete intervals of equal width, m, as in a histogram.

Third, for the r^{th} interval, the values of the two kernel densities (i.e., the heights of the densities) were computed at the midpoint of that interval, and stored. Call those values n_{r1} and n_{r2} for job levels 1 and 2, respectively, and observe that the two rectangles with areas $mn_{r1}$and mn_{r2} approximate, respectively, the areas under the two density functions within the r^{th} interval. The shorter rectangle is contained in the larger one and, therefore, captures the amount of overlap in the wage distributions within the r^{th} interval.

Fourth, the third step was repeated for every interval, and the results were summed across intervals. That is, $O_{1,2} = \sum_r \min\{mn_{r1}, mn_{r2}\}$, where the sum is over all intervals. So $O_{1,2}$ approximates the amount of overlap in the wage distributions between job levels 1 and 2; it lies in $[0, 1]$, with 1 representing complete overlap.

In the data, $O_{1,2}$ was found to be 0.339, which represents a considerable amount of overlap. According to the model of Gibbons and Waldman (2006), $O_{1,2}$ is positive because workers with different levels of education are pooled, and if $O_{1,2}$ were to be computed within a particular education group it should be 0. But in the data, the values of $O_{1,2}$ that are computed within educational groups are even larger than the preceding value that pools all observations. These values are 0.340 for high school graduates, 0.342 for college graduates, 0.414 for master's degree holders, and 0.368 for holders of doctoral degrees. Thus, in these data the wage overlap across job levels is not driven by pooling of workers with different education levels.

The present model offers a new theoretical explanation for the wage overlap observed in the preceding personnel data, and in other data. The explanation is based on the combination of multi-tasking and a positive probability that competing employers observe the workers' performances.

Definition 2 *Wage distributions are said to overlap across adjacent job levels if for some realizations of the random variables the worker assigned to job* 1 *receives a relatively higher period-*2 *wage than the worker in job* 2*, whereas for others the worker in job* 2 *receives a higher wage.*

Proposition 2 has the following corollary:

Corollary 2 *In the equilibrium described in Proposition 2:*
a) If $\min\{p_\alpha, p_\beta\} > 0$*, then wage distributions overlap across adjacent job levels.*
b) If $0 = \min\{p_\alpha, p_\beta\} < \max\{p_\alpha, p_\beta\}$*, then wage distributions could overlap across adjacent job levels, but not necessarily for all parameter configurations.*
c) If $p_\alpha = p_\beta = 0$*, then wage distributions do not overlap across adjacent job levels.*

Putt's Law states that workers are sometimes denied promotions precisely *because* they excel in their jobs; by rendering themselves indispensable in their current jobs they become "too good to promote".[40] The explanation for the overlap in wage distributions in the present study relates to the discussion of Putt's Law in DeVaro and Gürtler (2016a). As indicated by the law, very able workers may be denied promotion, and the same is true here. This obviously hurts these workers if external firms are unable to observe workers' performances, and wages in job 2 tend to be higher than those in job 1. However, if workers' performances are observable, external firms can infer workers' abilities, so that the very able workers receive high wages even if they are denied promotion.

7 Conclusions

The two pillars of the theoretical literature on careers (i.e., tournament models and learning-based job-assignment models) have stood largely in isolation, though recent work on (single-task) market-based tournaments has brought them closer together. Moreover, multi-task jobs are important real-world phenomena that have been neglected in both pillars of the literature and that have never been analyzed in an integrative, market-based model that treats incentives and assignment simultaneously under various assumptions about employer learning. This study provides a general, multi-task promotion model that incorporates both task-specific abilities and efforts and that nests symmetric and asymmetric learning (and a mix of both), thereby offering an organizing framework for interpreting the entire theoretical promotions literature.

A result of the general model is that strategic shirking occurs in multiple-firm labor markets with market-based wage setting, just as recent research has shown it to occur in a single-firm context with classic wage setting. The logic is similar in both settings. Worker behavior is driven by the wage spread (or anticipated spread) between job levels, regardless of the mechanism that generates that spread. Other results are unique products of the general integrative model and could not have been anticipated based on either of the pillars considered in isolation. For example, the result that strategic shirking extends to market-based settings is shown to hinge on the nature of the employer learning that occurs in the labor market. When learning is purely symmetric, strategic shirking does not occur on either task. In other words, in a market-based setting, asymmetric employer learning is a necessary condition for strategic shirking to occur. A growing body of empirical evidence suggests the importance of asymmetric learning in labor markets, particularly among college-educated workers, which, given the present results, implies strategic shirking.[41]

The model yields results that are not only unique but surprising and that overturn conventional wisdom. A natural intuition is that welfare is increased if the labor market transitions from asymmetric towards

[40] Putt (2006) contains anecdotal examples of the Law, along with the following definition of it: "*Technology is dominated by two types of people: those who understand what they do not manage and those who manage what they do not understand.*" Related concepts are the "Dilbert Principle" and "Putt's Corollary" (sometimes called the "First Corollary to Putt's Law"), which states that every technical hierarchy eventually develops a competence inversion, i.e., over time the incompetent workers get promoted to "manage what they do not understand" and the competent ones are left in lower-level jobs to "understand what they do not manage".

[41] See, e.g., Schönberg (2007), DeVaro and Waldman (2012), Kim and Usui (2014), Cassidy et al. (2016), and Fan and DeVaro (2020).

symmetric employer learning. That intuition derives from the job-assignment pillar, because tournament theory typically neglects efficiency in job assigments. The basis for the intuition (which requires either flexible managerial job slots or fixed slots where the managerial position can go unstaffed) is that asymmetric learning generates inefficient job assignments and that the inefficiency vanishes in the case of symmetric learning. Although efficiency in effort allocation is usually ignored in such models, it is shown here that when learning becomes more symmetric, there are ambiguous welfare consequences. Efficiency in job assignments improves (if managerial job slots are flexible, or fixed but can go unstaffed), but efficiency in effort allocation can either increase or decrease.

The preceding insight suggests policy implications both for the government and for individual employers. Technological change, or government policies designed to reduce inefficiencies in the labor market in the area of job assignments (by rendering employer learning more symmetric) may have the unintended consequence of worsening the allocation of worker effort. From the employer's perspective, a common approach is to implement some form of incentive compensation that complements and supplements the promotion-based incentives that naturally occur whenever a job hierarchy exists in conjunction with the possibility of internal promotion. The preceding results show that such compensation policies that are designed to increase effort could actually be welfare-reducing in some jobs (though welfare-enhancing in others) if the equilibrium effort levels in the absence of such pay schemes are at or above those of the full-information benchmark. These results add an important new item to the list of known pitfalls associated with incentive pay and strikingly highlight the important interplay between two distinct sources of incentives (i.e., promotions and performance-based pay) that employers often use simultaneously.

8 Appendix

Proofs of lemmas and propositions:

Proof of Proposition 1. Consider the efficient promotion rule. Given the assumption that exactly one of the workers must be promoted to job 2 in period 2, the only question is whether worker 1 or 2 is promoted. If worker 1 is promoted, total period-2 surplus is

$$(1+S)\left(c_2+d_{2\alpha}a_{1\alpha}+qe_{1\alpha}^2+d_{2\beta}a_{1\beta}+qe_{1\beta}^2+c_1+d_{1\alpha}a_{2\alpha}+qe_{2\alpha}^2+d_{1\beta}a_{2\beta}+qe_{2\beta}^2\right)$$
$$-\sum_{j\in\{\alpha,\beta\}}\kappa\left(e_{1j}^2\right)-\sum_{j\in\{\alpha,\beta\}}\kappa\left(e_{2j}^2\right).$$

In contrast, if worker 2 is promoted, total period-2 surplus is

$$(1+S)\left(c_1+d_{1\alpha}a_{1\alpha}+qe_{1\alpha}^2+d_{1\beta}a_{1\beta}+qe_{1\beta}^2+c_2+d_{2\alpha}a_{2\alpha}+qe_{2\alpha}^2+d_{2\beta}a_{2\beta}+qe_{2\beta}^2\right)$$
$$-\sum_{j\in\{\alpha,\beta\}}\kappa\left(e_{1j}^2\right)-\sum_{j\in\{\alpha,\beta\}}\kappa\left(e_{2j}^2\right).$$

The former surplus is higher than the latter if and only if

$$(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0.$$

Disregarding all terms that are independent of effort, the efficient effort by worker i on task α in period t is given by

$$e_{i\alpha}^{fb} \in \arg\max_{e_{i\alpha}^t \in \mathbb{R}} \left(\left(1 + s_{hi}^t\right) q \left(e_{i\alpha}^t + e_{i\beta}^t\right) - \sum_{j \in \{\alpha,\beta\}} \kappa \left(e_{ij}^t\right) \right).$$

Obviously, the efficient effort level is never negative. Assuming that this effort can be determined by the first-order condition to the surplus-maximization problem, the following expression obtains:

$$\left(1 + s_{hi}^t\right) q - \kappa_+ e_{i\alpha}^t = 0 \Leftrightarrow e_{i\alpha}^{fb} = \frac{\left(1 + s_{hi}^t\right) q}{\kappa_+}.$$

The same effort obtains on task β, so that the task-related subscript can be dropped. ■

Proof of Lemma 1. First, note that both workers choose the regular effort level of zero in period 2 because the game ends after that period. All firms correctly anticipate this behavior. If the period-1 employer rehires worker i in period 2, the worker's period-2 output would be at least

$$(1 + S) \cdot \min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\underline{a}\right\}.$$

In contrast, if one of the external firms hires the worker, the worker's period-2 output would be at most

$$c_2 + (d_{2\alpha} + d_{2\beta})\overline{a}.$$

Suppose that

$$(1 + S) \cdot \min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\underline{a}\right\} > c_2 + (d_{2\alpha} + d_{2\beta})\overline{a}$$
$$\Leftrightarrow S > \frac{c_2 + (d_{2\alpha} + d_{2\beta})\overline{a}}{\min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\underline{a}\right\}} - 1.$$

It is then easy to understand that no equilibrium exists in which one of the external firms hires the worker away from the period-1 employer (unless the employer mistakenly does not make a counteroffer). To show this, denote the highest wage offer from the external firms by w_{ef}. If we have $w_{ef} < (1 + S)\min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\underline{a}\right\}$, the period-1 employer reacts by matching w_{ef} and, thus, retains the worker. If $w_{ef} \geq (1 + S)\min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\underline{a}\right\}$, there are two possibilities. Either the period-1 employer reacts by matching w_{ef} and thus retains the worker, or the period-1 employer does not match the offer, and the external firm offering the highest wage hires the worker. Then, however, making a loss, this firm has an incentive to deviate to any lower wage offer (when several external firms make the same offer, they all have an incentive to deviate to a lower wage). This proves the lemma with

$$S_1 := \frac{c_2 + (d_{2\alpha} + d_{2\beta})\,\overline{a}}{\min\left\{c_1 + (d_{1\alpha} + d_{1\beta})\,\underline{a}, c_2 + (d_{2\alpha} + d_{2\beta})\,\underline{a}\right\}} - 1 > 0.$$

■

Proof of Proposition 2. As previously stated, both workers choose the regular effort level of zero in period 2 because the game ends after that period. The remainder of the proof proceeds in a number of steps. The general structure of period-2 wages is determined first (part i)). It is then assumed that the external firms and both workers believe that the incumbent firm uses the promotion rule specified in the proposition and that all firms believe that the two workers choose the same efforts in period 1. Building on these two assumptions, it is shown that the incumbent firm's period-2 wage bill does not depend on which of the two workers the firm promotes (part ii)). It is then shown that the incumbent firm finds it optimal to indeed use the promotion rule specified in the proposition (part iii)). In the next step, workers' optimal period-1 efforts are determined, and it is shown that both workers in fact choose the same efforts (part iv)). In the next step, specific expressions for the period-2 wages are determined (part v)).

i) It is first shown that, in general, the period-2 wage for worker i when assigned to job k_i by the incumbent firm, and given realizations of σ_α and σ_β, can be stated as

$$w_{i2}\left(\sigma_\alpha, \sigma_\beta, k_i\right) = c_2 + d_{2\alpha}\tilde{a}^E_{i\alpha}\left(\sigma_\alpha, \sigma_\beta, k_i\right) + d_{2\beta}\tilde{a}^E_{i\beta}\left(\sigma_\alpha, \sigma_\beta, k_i\right).$$

Because $S \geq S_1$, the external firms can hire worker i only if the incumbent firm mistakenly fails to make a counteroffer. Since those mistakes do not depend on the worker's ability, the expected ability of workers that are actually switching firms is equal to the overall expected ability of workers, which the external firms believe to be equal to $\tilde{a}^E_{i\alpha}\left(\sigma_\alpha, \sigma_\beta, k_i\right)$ and $\tilde{a}^E_{i\beta}\left(\sigma_\alpha, \sigma_\beta, k_i\right)$, respectively. For the external firms, the expected period-2 output of a worker actually switching firms can thus be determined as

$$Z := c_2 + d_{2\alpha}\tilde{a}^E_{i\alpha}\left(\sigma_\alpha, \sigma_\beta, k_i\right) + d_{2\beta}\tilde{a}^E_{i\beta}\left(\sigma_\alpha, \sigma_\beta, k_i\right),$$

because a newly hired worker is always assigned to job 2. Notice that because the incumbent firm sometimes mistakenly fails to make a counteroffer, the external firms' highest equilibrium wage offer w_{ef} always satisfies $w_{ef} \leq Z$ (since at least one external firm would expect to make a loss otherwise). Unless it mistakenly fails to make a counteroffer, the period-1 employer will always match w_{ef}. This means that the worker's period-2 wage is always given by w_{ef} (regardless of whether the period-1 employer manages or fails to make a counteroffer). Since the external firms compete in a regular Bertrand fashion, it is straightforward to see that $w_{ef} = Z$. Hence,

$$w_{i2}\left(\sigma_\alpha, \sigma_\beta, k_i\right) = w_{ef} = Z.$$

ii) As previously indicated, it is assumed that the external firms and both workers believe that the incumbent firm promotes worker 1 to job 2 at the beginning of period 2 if and only if $(d_{2\alpha} - d_{1\alpha})\left(\tilde{a}^I_{1\alpha} - \tilde{a}^I_{2\alpha}\right) + (d_{1\beta} - d_{2\beta})\left(\tilde{a}^I_{2\beta} - \tilde{a}^I_{1\beta}\right) \geq 0$, and it is further assumed that $\tilde{e}^1_{1j} = \tilde{e}^1_{2j}$ for all j. To simplify notation, define $\mathbf{y} := \left(y^{\alpha}_{111}, y^{\beta}_{111}, y^{\alpha}_{211}, y^{\beta}_{211}\right)$. The objective of this part of the proof is to show that the incumbent firm's expected period-2 wage bill does not depend on which of the two workers the firm promotes. Defining

$$W_{i2}(k_i) := p_{\alpha\beta} w_{i2}(\alpha, \beta, k_i) + p_{\alpha\neg\beta} w_{i2}(\alpha, \neg\beta, k_i) + p_{\neg\alpha\beta} w_{i2}(\neg\alpha, \beta, k_i) + p_{\neg\alpha\neg\beta} w_{i2}(\neg\alpha, \neg\beta, k_i),$$

the objective is, thus, to show that

$$E\left[W_{12}(2) + W_{22}(1) \middle| \mathbf{y}\right] = E\left[W_{12}(1) + W_{22}(2) \middle| \mathbf{y}\right]$$

for all $\mathbf{y}$. The condition is shown to be true by verifying that

$$E\left[w_{12}(\sigma_\alpha, \sigma_\beta, 2) + w_{22}(\sigma_\alpha, \sigma_\beta, 1) \middle| \mathbf{y}\right] = E\left[w_{12}(\sigma_\alpha, \sigma_\beta, 1) + w_{22}(\sigma_\alpha, \sigma_\beta, 2) \middle| \mathbf{y}\right]$$

for all $\sigma_\alpha, \sigma_\beta, \mathbf{y}$. Using

$$w_{i2}(\sigma_\alpha, \sigma_\beta, k_i) = c_2 + d_{2\alpha}\tilde{a}^E_{i\alpha}(\sigma_\alpha, \sigma_\beta, k_i) + d_{2\beta}\tilde{a}^E_{i\beta}(\sigma_\alpha, \sigma_\beta, k_i),$$

the last condition can be restated as

$$\begin{aligned} &E\left[d_{2\alpha}\tilde{a}^E_{1\alpha}(\sigma_\alpha, \sigma_\beta, 2) + d_{2\beta}\tilde{a}^E_{1\beta}(\sigma_\alpha, \sigma_\beta, 2) + d_{2\alpha}\tilde{a}^E_{2\alpha}(\sigma_\alpha, \sigma_\beta, 1) + d_{2\beta}\tilde{a}^E_{2\beta}(\sigma_\alpha, \sigma_\beta, 1) \middle| \mathbf{y}\right] = \\ &E\left[d_{2\alpha}\tilde{a}^E_{1\alpha}(\sigma_\alpha, \sigma_\beta, 1) + d_{2\beta}\tilde{a}^E_{1\beta}(\sigma_\alpha, \sigma_\beta, 1) + d_{2\alpha}\tilde{a}^E_{2\alpha}(\sigma_\alpha, \sigma_\beta, 2) + d_{2\beta}\tilde{a}^E_{2\beta}(\sigma_\alpha, \sigma_\beta, 2) \middle| \mathbf{y}\right]. \end{aligned}$$

Define $\tilde{\Phi}^{IE}(\sigma_\alpha, \sigma_\beta) := (d_{2\alpha} - d_{1\alpha})\left(\tilde{a}^{IE}_{2\alpha}(\sigma_\alpha) - \tilde{a}^{IE}_{1\alpha}(\sigma_\alpha)\right) + (d_{1\beta} - d_{2\beta})\left(\tilde{a}^{IE}_{1\beta}(\sigma_\beta) - \tilde{a}^{IE}_{2\beta}(\sigma_\beta)\right)$. If $\sigma_\alpha = \alpha$ and $\sigma_\beta = \beta$, then the condition can be further rewritten as

$$\begin{aligned} &E\left[E\left[d_{2\alpha}\frac{y^{\alpha}_{111} - q\tilde{e}^1_{1\alpha}}{d_{1\alpha}} + d_{2\beta}\frac{y^{\beta}_{111} - q\tilde{e}^1_{1\beta}}{d_{1\beta}} + d_{2\alpha}\frac{y^{\alpha}_{211} - q\tilde{e}^1_{2\alpha}}{d_{1\alpha}} + d_{2\beta}\frac{y^{\beta}_{211} - q\tilde{e}^1_{2\beta}}{d_{1\beta}} \middle| \tilde{\Phi}^{IE}(\alpha, \beta) \leq 0\right] \middle| \mathbf{y}\right] = \\ &E\left[E\left[d_{2\alpha}\frac{y^{\alpha}_{111} - q\tilde{e}^1_{1\alpha}}{d_{1\alpha}} + d_{2\beta}\frac{y^{\beta}_{111} - q\tilde{e}^1_{1\beta}}{d_{1\beta}} + d_{2\alpha}\frac{y^{\alpha}_{211} - q\tilde{e}^1_{2\alpha}}{d_{1\alpha}} + d_{2\beta}\frac{y^{\beta}_{211} - q\tilde{e}^1_{2\beta}}{d_{1\beta}} \middle| \tilde{\Phi}^{IE}(\alpha, \beta) \geq 0\right] \middle| \mathbf{y}\right], \end{aligned}$$

which is true because

$$E\left[d_{2\alpha}\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{111}^{\beta}-q\tilde{e}_{1\beta}^{1}}{d_{1\beta}}+d_{2\alpha}\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{211}^{\beta}-q\tilde{e}_{2\beta}^{1}}{d_{1\beta}}\right|\left.\tilde{\Phi}^{IE}\left(\alpha,\beta\right)\leq 0\right]=$$

$$E\left[d_{2\alpha}\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{111}^{\beta}-q\tilde{e}_{1\beta}^{1}}{d_{1\beta}}+d_{2\alpha}\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{211}^{\beta}-q\tilde{e}_{2\beta}^{1}}{d_{1\beta}}\right|\left.\tilde{\Phi}^{IE}\left(\alpha,\beta\right)\geq 0\right]$$

$$=d_{2\alpha}\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{111}^{\beta}-q\tilde{e}_{1\beta}^{1}}{d_{1\beta}}+d_{2\alpha}\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\frac{y_{211}^{\beta}-q\tilde{e}_{2\beta}^{1}}{d_{1\beta}}.$$

Suppose now that $\sigma_{\alpha}=\alpha$ and $\sigma_{\beta}=\neg\beta$. Then it must be shown that

$$E\left[E\left[d_{2\alpha}\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\tilde{a}_{1\beta}^{E}\left(\alpha,\neg\beta,2\right)+d_{2\alpha}\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\tilde{a}_{2\beta}^{E}\left(\alpha,\neg\beta,1\right)\right|\tilde{\Phi}^{IE}\left(\alpha,\neg\beta\right)\leq 0\right]\left|\,\mathbf{y}\right]=$$

$$E\left[E\left[d_{2\alpha}\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\tilde{a}_{1\beta}^{E}\left(\alpha,\neg\beta,1\right)+d_{2\alpha}\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}+d_{2\beta}\tilde{a}_{2\beta}^{E}\left(\alpha,\neg\beta,2\right)\right|\tilde{\Phi}^{IE}\left(\alpha,\neg\beta\right)\geq 0\right]\left|\,\mathbf{y}\right],$$

or, equivalently,

$$E\left[E\left[\tilde{a}_{1\beta}^{E}\left(\alpha,\neg\beta,2\right)+\tilde{a}_{2\beta}^{E}\left(\alpha,\neg\beta,1\right)\right|\tilde{\Phi}^{IE}\left(\alpha,\neg\beta\right)\leq 0\right]\left|\,\mathbf{y}\right]=$$

$$E\left[E\left[\tilde{a}_{1\beta}^{E}\left(\alpha,\neg\beta,1\right)+\tilde{a}_{2\beta}^{E}\left(\alpha,\neg\beta,2\right)\right|\tilde{\Phi}^{IE}\left(\alpha,\neg\beta\right)\geq 0\right]\left|\,\mathbf{y}\right].$$

The latter condition is fulfilled if

$$E\left[a_{1\beta}+a_{2\beta}\right|a_{2\beta}-a_{1\beta}\geq\frac{d_{2\alpha}-d_{1\alpha}}{d_{1\beta}-d_{2\beta}}\left.\left(\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}-\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}\right)\right]=$$

$$E\left[a_{1\beta}+a_{2\beta}\right|a_{2\beta}-a_{1\beta}\leq\frac{d_{2\alpha}-d_{1\alpha}}{d_{1\beta}-d_{2\beta}}\left.\left(\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}-\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}\right)\right].$$

Express $a_{i\beta}$ as $a_{i\beta}=E\left[a\right]+\varepsilon_{i\beta}$, where $\varepsilon_{i\beta}$ is a random variable that has a symmetric distribution around zero, and define $\xi:=\frac{d_{2\alpha}-d_{1\alpha}}{d_{1\beta}-d_{2\beta}}\left(\frac{y_{211}^{\alpha}-q\tilde{e}_{2\alpha}^{1}}{d_{1\alpha}}-\frac{y_{111}^{\alpha}-q\tilde{e}_{1\alpha}^{1}}{d_{1\alpha}}\right)$. It then remains to show that

$$E\left[\varepsilon_{1\beta}+\varepsilon_{2\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\geq\xi\right]=$$

$$E\left[\varepsilon_{1\beta}+\varepsilon_{2\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\leq\xi\right].$$

It is easy to show that this equality is true, because both terms are equal to zero. To see this, notice that

$$E\left[\varepsilon_{1\beta}+\varepsilon_{2\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\geq(\leq)\,\xi\right]=$$

$$E\left[\varepsilon_{1\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\geq(\leq)\,\xi\right]+E\left[\varepsilon_{2\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\geq(\leq)\,\xi\right]=$$

$$E\left[\varepsilon_{1\beta}\right|\left.\varepsilon_{2\beta}-\varepsilon_{1\beta}\geq(\leq)\,\xi\right]-E\left[\varepsilon_{1\beta}\right|\left.-\varepsilon_{1\beta}+\varepsilon_{2\beta}\geq(\leq)\,\xi\right]=0,$$

because $\varepsilon_{2\beta}$ and $-\varepsilon_{1\beta}$ have the same distribution.

The proof is analogous in the case $\sigma_\alpha = \neg\alpha$ and $\sigma_\beta = \beta$, so it remains to consider the case $\sigma_\alpha = \neg\alpha$ and $\sigma_\beta = \neg\beta$. In this final case, the condition

$$E\left[\left. d_{2\alpha}\tilde{a}^E_{1\alpha}(\sigma_\alpha,\sigma_\beta,2) + d_{2\beta}\tilde{a}^E_{1\beta}(\sigma_\alpha,\sigma_\beta,2) + d_{2\alpha}\tilde{a}^E_{2\alpha}(\sigma_\alpha,\sigma_\beta,1) + d_{2\beta}\tilde{a}^E_{2\beta}(\sigma_\alpha,\sigma_\beta,1)\right| \mathbf{y}\right] =$$
$$E\left[\left. d_{2\alpha}\tilde{a}^E_{1\alpha}(\sigma_\alpha,\sigma_\beta,1) + d_{2\beta}\tilde{a}^E_{1\beta}(\sigma_\alpha,\sigma_\beta,1) + d_{2\alpha}\tilde{a}^E_{2\alpha}(\sigma_\alpha,\sigma_\beta,2) + d_{2\beta}\tilde{a}^E_{2\beta}(\sigma_\alpha,\sigma_\beta,2)\right| \mathbf{y}\right]$$

becomes

$$E\left[d_{2\alpha}a_{1\alpha} + d_{2\beta}a_{1\beta} + d_{2\alpha}a_{2\alpha} + d_{2\beta}a_{2\beta} \right| (d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0] =$$
$$E\left[d_{2\alpha}a_{1\alpha} + d_{2\beta}a_{1\beta} + d_{2\alpha}a_{2\alpha} + d_{2\beta}a_{2\beta} \right| (d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \leq 0].$$

Since $a_{1\alpha}$ and $a_{2\alpha}$ are i.i.d. and the same is true for $a_{1\beta}$ and $a_{2\beta}$, the conditional expectation on the right-hand-side of the condition can be stated as

$$\begin{aligned} & E\left[d_{2\alpha}a_{1\alpha} + d_{2\beta}a_{1\beta} + d_{2\alpha}a_{2\alpha} + d_{2\beta}a_{2\beta} \right| (d_{2\alpha} - d_{1\alpha})(a_{2\alpha} - a_{1\alpha}) + (d_{1\beta} - d_{2\beta})(a_{1\beta} - a_{2\beta}) \leq 0] \\ = \quad & E\left[d_{2\alpha}a_{1\alpha} + d_{2\beta}a_{1\beta} + d_{2\alpha}a_{2\alpha} + d_{2\beta}a_{2\beta} \right| (d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0], \end{aligned}$$

which completes this part of the proof.

iii) Turning now to the incumbent firm's choice of promotion rule, when the firm assigns workers 1 and 2 to jobs k_1 and k_2, expected period-2 profit can be stated as

$$\begin{aligned} E[\pi(k_1,k_2)|\,\mathbf{y}] \quad = \quad & (1-\gamma)(1+S)\left(c_{k_1} + d_{k_1\alpha}\tilde{a}^I_{1\alpha} + d_{k_1\beta}\tilde{a}^I_{1\beta}\right) - (1-\gamma)E[W_{12}(k_1)|\,\mathbf{y}] \\ & + (1-\gamma)(1+S)\left(c_{k_2} + d_{k_2\alpha}\tilde{a}^I_{2\alpha} + d_{k_2\beta}\tilde{a}^I_{2\beta}\right) - (1-\gamma)E[W_{22}(k_2)|\,\mathbf{y}]. \end{aligned}$$

The firm promotes worker 1 to job 2 if and only if

$$E[\pi(2,1)|\,\mathbf{y}] \geq E[\pi(1,2)|\,\mathbf{y}].$$

This condition is equivalent to

$$\begin{aligned} & (1-\gamma)(1+S)\left(c_2 + d_{2\alpha}\tilde{a}^I_{1\alpha} + d_{2\beta}\tilde{a}^I_{1\beta}\right) - (1-\gamma)E[W_{12}(2)|\,\mathbf{y}] \\ & + (1-\gamma)(1+S)\left(c_1 + d_{1\alpha}\tilde{a}^I_{2\alpha} + d_{1\beta}\tilde{a}^I_{2\beta}\right) - (1-\gamma)E[W_{22}(1)|\,\mathbf{y}] \\ & \geq (1-\gamma)(1+S)\left(c_1 + d_{1\alpha}\tilde{a}^I_{1\alpha} + d_{1\beta}\tilde{a}^I_{1\beta}\right) - (1-\gamma)E[W_{12}(1)|\,\mathbf{y}] \\ & + (1-\gamma)(1+S)\left(c_2 + d_{2\alpha}\tilde{a}^I_{2\alpha} + d_{2\beta}\tilde{a}^I_{2\beta}\right) - (1-\gamma)E[W_{22}(2)|\,\mathbf{y}] \end{aligned}$$

$$\Leftrightarrow \quad d_{2\alpha}\tilde{a}^{I}_{1\alpha} + d_{2\beta}\tilde{a}^{I}_{1\beta} + d_{1\alpha}\tilde{a}^{I}_{2\alpha} + d_{1\beta}\tilde{a}^{I}_{2\beta} \geq d_{1\alpha}\tilde{a}^{I}_{1\alpha} + d_{1\beta}\tilde{a}^{I}_{1\beta} + d_{2\alpha}\tilde{a}^{I}_{2\alpha} + d_{2\beta}\tilde{a}^{I}_{2\beta}$$

$$\Leftrightarrow \quad (d_{2\alpha} - d_{1\alpha})\left(\tilde{a}^{I}_{1\alpha} - \tilde{a}^{I}_{2\alpha}\right) + (d_{1\beta} - d_{2\beta})\left(\tilde{a}^{I}_{2\beta} - \tilde{a}^{I}_{1\beta}\right) \geq 0,$$

where the condition $E\left[W_{12}(2)|\,\mathbf{y}\right] + E\left[W_{22}(1)|\,\mathbf{y}\right] = E\left[W_{12}(1)|\,\mathbf{y}\right] + E\left[W_{22}(2)|\,\mathbf{y}\right]$ is used in the transformations of the inequality.

iv) In the next step, the two workers' optimal choices of period-1 efforts are determined. As mentioned before, it is assumed that both workers believe that the incumbent firm promotes worker 1 to job 2 at the beginning of period 2 if and only if $(d_{2\alpha} - d_{1\alpha})\left(\tilde{a}^{I}_{1\alpha} - \tilde{a}^{I}_{2\alpha}\right) + (d_{1\beta} - d_{2\beta})\left(\tilde{a}^{I}_{2\beta} - \tilde{a}^{I}_{1\beta}\right) \geq 0$. This condition is equivalent to

$$(d_{2\alpha} - d_{1\alpha})\left(\frac{y^{\alpha}_{111} - q\tilde{e}^{1}_{1\alpha}}{d_{1\alpha}} - \frac{y^{\alpha}_{211} - q\tilde{e}^{1}_{2\alpha}}{d_{1\alpha}}\right) + (d_{1\beta} - d_{2\beta})\left(\frac{y^{\beta}_{211} - q\tilde{e}^{1}_{2\beta}}{d_{1\beta}} - \frac{y^{\beta}_{111} - q\tilde{e}^{1}_{1\beta}}{d_{1\beta}}\right) \geq 0$$

$$\Leftrightarrow (d_{2\alpha} - d_{1\alpha})\left(a_{1\alpha} - a_{2\alpha} + \frac{q\left(e^{1}_{1\alpha} - \tilde{e}^{1}_{1\alpha} - e^{1}_{2\alpha} + \tilde{e}^{1}_{2\alpha}\right)}{d_{1\alpha}}\right)$$

$$+ (d_{1\beta} - d_{2\beta})\left(a_{2\beta} - a_{1\beta} + \frac{q\left(e^{1}_{2\beta} - \tilde{e}^{1}_{2\beta} - e^{1}_{1\beta} + \tilde{e}^{1}_{1\beta}\right)}{d_{1\beta}}\right) \geq 0$$

$$\Leftrightarrow (d_{2\alpha} - d_{1\alpha})(a_{2\alpha} - a_{1\alpha}) + (d_{1\beta} - d_{2\beta})(a_{1\beta} - a_{2\beta})$$

$$\leq (d_{2\alpha} - d_{1\alpha})\left(\frac{q\left(e^{1}_{1\alpha} - \tilde{e}^{1}_{1\alpha} - e^{1}_{2\alpha} + \tilde{e}^{1}_{2\alpha}\right)}{d_{1\alpha}}\right) + (d_{1\beta} - d_{2\beta})\left(\frac{q\left(e^{1}_{2\beta} - \tilde{e}^{1}_{2\beta} - e^{1}_{1\beta} + \tilde{e}^{1}_{1\beta}\right)}{d_{1\beta}}\right).$$

Using the previous definitions for Φ, G and $W_{i2}(k_i)$, worker 1's period-1 optimization problem can be stated as

$$\begin{aligned}
&Max_{\left(e^{1}_{1\alpha}, e^{1}_{1\beta}\right)\in\mathbb{R}^2}\\
&G\left((d_{2\alpha} - d_{1\alpha})\left(\frac{q\left(e^{1}_{1\alpha} - \tilde{e}^{1}_{1\alpha} - e^{1}_{2\alpha} + \tilde{e}^{1}_{2\alpha}\right)}{d_{1\alpha}}\right) + (d_{1\beta} - d_{2\beta})\left(\frac{q\left(e^{1}_{2\beta} - \tilde{e}^{1}_{2\beta} - e^{1}_{1\beta} + \tilde{e}^{1}_{1\beta}\right)}{d_{1\beta}}\right)\right) E_{\mathbf{a}}\left[W_{12}(2)\right]\\
&+\left(1 - G\left((d_{2\alpha} - d_{1\alpha})\left(\frac{q\left(e^{1}_{1\alpha} - \tilde{e}^{1}_{1\alpha} - e^{1}_{2\alpha} + \tilde{e}^{1}_{2\alpha}\right)}{d_{1\alpha}}\right) + (d_{1\beta} - d_{2\beta})\left(\frac{q\left(e^{1}_{2\beta} - \tilde{e}^{1}_{2\beta} - e^{1}_{1\beta} + \tilde{e}^{1}_{1\beta}\right)}{d_{1\beta}}\right)\right)\right)\cdot\\
&E_{\mathbf{a}}\left[W_{12}(1)\right] - \sum_{j}\kappa\left(e^{1}_{1j}\right).
\end{aligned}$$

Worker 2's optimization problem is

$$Max_{\left(e^1_{2\alpha},e^1_{2\beta}\right)\in\mathbb{R}^2}$$
$$G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e^1_{1\alpha}-\tilde{e}^1_{1\alpha}-e^1_{2\alpha}+\tilde{e}^1_{2\alpha}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e^1_{2\beta}-\tilde{e}^1_{2\beta}-e^1_{1\beta}+\tilde{e}^1_{1\beta}\right)}{d_{1\beta}}\right)\right)E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]$$
$$+\left(1-G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e^1_{1\alpha}-\tilde{e}^1_{1\alpha}-e^1_{2\alpha}+\tilde{e}^1_{2\alpha}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e^1_{2\beta}-\tilde{e}^1_{2\beta}-e^1_{1\beta}+\tilde{e}^1_{1\beta}\right)}{d_{1\beta}}\right)\right)\right)\cdot$$
$$E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-\sum_j\kappa\left(e^1_{2j}\right).$$

Optimal efforts are assumed to be characterized by the following first-order conditions to the preceding maximization problems:

$$G'\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e^1_{1\alpha}-\tilde{e}^1_{1\alpha}-e^1_{2\alpha}+\tilde{e}^1_{2\alpha}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e^1_{2\beta}-\tilde{e}^1_{2\beta}-e^1_{1\beta}+\tilde{e}^1_{1\beta}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)$$
$$+G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e^1_{1\alpha}-\tilde{e}^1_{1\alpha}-e^1_{2\alpha}+\tilde{e}^1_{2\alpha}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e^1_{2\beta}-\tilde{e}^1_{2\beta}-e^1_{1\beta}+\tilde{e}^1_{1\beta}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e^1_{1\alpha}}$$
$$+\left(1-G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e^1_{1\alpha}-\tilde{e}^1_{1\alpha}-e^1_{2\alpha}+\tilde{e}^1_{2\alpha}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e^1_{2\beta}-\tilde{e}^1_{2\beta}-e^1_{1\beta}+\tilde{e}^1_{1\beta}\right)}{d_{1\beta}}\right)\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e^1_{1\alpha}}=\kappa'\left(e^1_{1\alpha}\right),$$

$$-G'\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)$$
$$+G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e_{1\beta}^{1}}$$
$$+\left(1-G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e_{1\beta}^{1}}=\kappa'\left(e_{1\beta}^{1}\right),$$

$$G'\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)$$
$$+\left(1-G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\alpha}^{1}}$$
$$+G\left(\left(d_{2\alpha}-d_{1\alpha}\right)\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+\left(d_{1\beta}-d_{2\beta}\right)\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\alpha}^{1}}=\kappa'\left(e_{2\alpha}^{1}\right),$$

$$-G'\left((d_{2\alpha}-d_{1\alpha})\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+(d_{1\beta}-d_{2\beta})\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)$$
$$+\left(1-G\left((d_{2\alpha}-d_{1\alpha})\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+(d_{1\beta}-d_{2\beta})\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\beta}^{1}}$$
$$+G\left((d_{2\alpha}-d_{1\alpha})\left(\frac{q\left(e_{1\alpha}^{1}-\tilde{e}_{1\alpha}^{1}-e_{2\alpha}^{1}+\tilde{e}_{2\alpha}^{1}\right)}{d_{1\alpha}}\right)+(d_{1\beta}-d_{2\beta})\left(\frac{q\left(e_{2\beta}^{1}-\tilde{e}_{2\beta}^{1}-e_{1\beta}^{1}+\tilde{e}_{1\beta}^{1}\right)}{d_{1\beta}}\right)\right)\cdot$$
$$\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\beta}^{1}}=\kappa'\left(e_{2\beta}^{1}\right).$$

In equilibrium, the firms correctly anticipate workers' efforts, thus $e_{ij}^{1}=\tilde{e}_{ij}^{1}$ for all i,j, implying that the above conditions simplify to

$$G'\left(0\right)\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)+G\left(0\right)\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e_{1\alpha}^{1}}+\left(1-G\left(0\right)\right)\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e_{1\alpha}^{1}}=\kappa'\left(e_{1\alpha}^{1}\right),$$

$$-G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)+G\left(0\right)\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e_{1\beta}^{1}}+\left(1-G\left(0\right)\right)\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e_{1\beta}^{1}}=\kappa'\left(e_{1\beta}^{1}\right),$$

$$G'\left(0\right)\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)+\left(1-G\left(0\right)\right)\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\alpha}^{1}}+G\left(0\right)\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\alpha}^{1}}=\kappa'\left(e_{2\alpha}^{1}\right),$$

$$-G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)+\left(1-G\left(0\right)\right)\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\beta}^{1}}+G\left(0\right)\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\beta}^{1}}=\kappa'\left(e_{2\beta}^{1}\right).$$

Since $G\left(0\right)=0.5$,

$$G'\left(0\right)\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)+0.5\left(\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e_{1\alpha}^{1}}+\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e_{1\alpha}^{1}}\right)=\kappa'\left(e_{1\alpha}^{1}\right),$$

$$-G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)+0.5\left(\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]}{\partial e_{1\beta}^{1}}+\frac{\partial E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]}{\partial e_{1\beta}^{1}}\right)=\kappa'\left(e_{1\beta}^{1}\right),$$

$$G'\left(0\right)\frac{q\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)+0.5\left(\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\alpha}^{1}}+\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\alpha}^{1}}\right)=\kappa'\left(e_{2\alpha}^{1}\right),$$

$$-G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]\right)+0.5\left(\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(2\right)\right]}{\partial e_{2\beta}^{1}}+\frac{\partial E_{\mathbf{a}}\left[W_{22}\left(1\right)\right]}{\partial e_{2\beta}^{1}}\right)=\kappa'\left(e_{2\beta}^{1}\right).$$

It is shown in part ii) of this proof that

$$E\left[W_{12}(2)|\,\mathbf{y}\right]+E\left[W_{22}(1)|\,\mathbf{y}\right]=E\left[W_{12}(1)|\,\mathbf{y}\right]+E\left[W_{22}(2)|\,\mathbf{y}\right]$$
$$\Leftrightarrow E\left[W_{12}(2)-W_{12}(1)|\,\mathbf{y}\right]=E\left[W_{22}(2)-W_{22}(1)|\,\mathbf{y}\right]$$

for all $\mathbf{y}$. The latter condition implies

$$E_{\mathbf{a}}\left[W_{12}(2)\right]-E_{\mathbf{a}}\left[W_{12}(1)\right]=E_{\mathbf{a}}\left[W_{22}(2)\right]-E_{\mathbf{a}}\left[W_{22}(1)\right].$$

Given the external firms' belief regarding the incumbent firm's promotion rule, in the eight possible scenarios for the information external employers receive concerning ability on a particular task, $\tilde{a}^E_{i\alpha}(\sigma_\alpha,\sigma_\beta,k_i)$ and $\tilde{a}^E_{i\beta}(\sigma_\alpha,\sigma_\beta,k_i)$ can be stated as

$$\tilde{a}^E_{i\alpha}(\alpha,\beta,1)=\tilde{a}^E_{i\alpha}(\alpha,\beta,2)=\tilde{a}^E_{i\alpha}(\alpha,\neg\beta,1)=\tilde{a}^E_{i\alpha}(\alpha,\neg\beta,2)=\frac{y^\alpha_{i11}-q\tilde{e}^1_{i\alpha}}{d_{1\alpha}},$$
$$\tilde{a}^E_{i\beta}(\alpha,\beta,1)=\tilde{a}^E_{i\beta}(\alpha,\beta,2)=\tilde{a}^E_{i\beta}(\neg\alpha,\beta,1)=\tilde{a}^E_{i\beta}(\neg\alpha,\beta,2)=\frac{y^\beta_{i11}-q\tilde{e}^1_{i\beta}}{d_{1\beta}},$$
$$\tilde{a}^E_{i\alpha}(\neg\alpha,\beta,2)=E_{a_{1\alpha},a_{2\alpha}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[a_{l\alpha}+\frac{1}{\Omega}\left(\frac{y^\beta_{i11}-y^\beta_{l11}-q\tilde{e}^1_{i\beta}+q\tilde{e}^1_{l\beta}}{d_{1\beta}}\right),\bar{a}\right]\right],$$
$$\tilde{a}^E_{i\alpha}(\neg\alpha,\beta,1)=E_{a_{1\alpha},a_{2\alpha}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}\left(\frac{y^\beta_{i11}-y^\beta_{l11}-q\tilde{e}^1_{i\beta}+q\tilde{e}^1_{l\beta}}{d_{1\beta}}\right)\right)\right],$$
$$\tilde{a}^E_{i\alpha}(\neg\alpha,\neg\beta,2)=E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[a_{l\alpha}+\frac{1}{\Omega}\left(a_{i\beta}-a_{l\beta}\right),\bar{a}\right]\right],$$
$$\tilde{a}^E_{i\alpha}(\neg\alpha,\neg\beta,1)=E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}\left(a_{i\beta}-a_{l\beta}\right)\right)\right],$$
$$\tilde{a}^E_{i\beta}(\alpha,\neg\beta,2)=E_{a_{1\beta},a_{2\beta}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega\left(\frac{y^\alpha_{i11}-y^\alpha_{l11}-q\left(\tilde{e}^1_{i\alpha}-\tilde{e}^1_{l\alpha}\right)}{d_{1\alpha}}\right)\right]\right],$$
$$\tilde{a}^E_{i\beta}(\alpha,\neg\beta,1)=E_{a_{1\beta},a_{2\beta}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega\left(\frac{y^\alpha_{i11}-y^\alpha_{l11}-q\left(\tilde{e}^1_{i\alpha}-\tilde{e}^1_{l\alpha}\right)}{d_{1\alpha}}\right),\bar{a}\right]\right],$$
$$\tilde{a}^E_{i\beta}(\neg\alpha,\neg\beta,2)=E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}\right)\right]\right],\text{ and}$$
$$\tilde{a}^E_{i\beta}(\neg\alpha,\neg\beta,1)=E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}\right),\bar{a}\right]\right],$$

with $i,l\in\{1,2\}$, $i\neq l$. To understand the formulas, take the expression for $\tilde{a}^E_{i\alpha}(\neg\alpha,\beta,2)$ as an example. That expression is the external firms' belief regarding worker i's ability on task α, given that performance on that task was not observable, performance on task β was observable, and worker i was promoted to job 2. Given the promotion of worker i and given the external firms' belief regarding the incumbent firm's promotion rule, the external firms believe that

$$(d_{2\alpha}-d_{1\alpha})\left(\tilde{a}_{i\alpha}^{IE}(\neg\alpha)-\tilde{a}_{l\alpha}^{IE}(\neg\alpha)\right)+(d_{1\beta}-d_{2\beta})\left(\tilde{a}_{l\beta}^{IE}(\beta)-\tilde{a}_{i\beta}^{IE}(\beta)\right)\geq 0$$
$$\Leftrightarrow \tilde{a}_{i\alpha}^{IE}(\neg\alpha)\geq \tilde{a}_{l\alpha}^{IE}(\neg\alpha)+\frac{1}{\Omega}\left(\tilde{a}_{i\beta}^{IE}(\beta)-\tilde{a}_{l\beta}^{IE}(\beta)\right).$$

Inserting $\frac{y_{i11}^{\beta}-q\tilde{e}_{i\beta}^{1}}{d_{1\beta}}$ for $\tilde{a}_{i\beta}^{IE}(\beta)$, $\frac{y_{l11}^{\beta}-q\tilde{e}_{l\beta}^{1}}{d_{1\beta}}$ for $\tilde{a}_{l\beta}^{IE}(\beta)$, $a_{i\alpha}$ for $\tilde{a}_{i\alpha}^{IE}(\neg\alpha)$, and $a_{l\alpha}$ for $\tilde{a}_{l\alpha}^{IE}(\neg\alpha)$, the condition becomes

$$a_{i\alpha}\geq a_{l\alpha}+\frac{1}{\Omega}\left(\frac{y_{i11}^{\beta}-y_{l11}^{\beta}-q\tilde{e}_{i\beta}^{1}+q\tilde{e}_{l\beta}^{1}}{d_{1\beta}}\right),$$

which gives the conditional expectation presented before. All other formulas can be derived analogously. Given that the formulas are symmetric for the two workers 1 and 2, it follows that

$$\begin{aligned}\frac{\partial\tilde{a}_{1\alpha}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{1\alpha}^{1}} &= \frac{\partial\tilde{a}_{2\alpha}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{2\alpha}^{1}},\frac{\partial\tilde{a}_{1\alpha}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{1\beta}^{1}}=\frac{\partial\tilde{a}_{2\alpha}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{2\beta}^{1}},\\ \frac{\partial\tilde{a}_{1\beta}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{1\alpha}^{1}} &= \frac{\partial\tilde{a}_{2\beta}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{2\alpha}^{1}}\text{ and }\frac{\partial\tilde{a}_{1\beta}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{1\beta}^{1}}=\frac{\partial\tilde{a}_{2\beta}^{E}(\sigma_\alpha,\sigma_\beta,k)}{\partial e_{2\beta}^{1}},\end{aligned}$$

for all $\sigma_\alpha,\sigma_\beta$ and $k\in\{1,2\}$.

Recall that

$$W_{i2}(k_i)=p_{\alpha\beta}w_{i2}(\alpha,\beta,k_i)+p_{\alpha\neg\beta}w_{i2}(\alpha,\neg\beta,k_i)+p_{\neg\alpha\beta}w_{i2}(\neg\alpha,\beta,k_i)+p_{\neg\alpha\neg\beta}w_{i2}(\neg\alpha,\neg\beta,k_i),$$

with

$$w_{i2}(\sigma_\alpha,\sigma_\beta,k_i)=c_2+d_{2\alpha}\tilde{a}_{i\alpha}^{E}(\sigma_\alpha,\sigma_\beta,k_i)+d_{2\beta}\tilde{a}_{i\beta}^{E}(\sigma_\alpha,\sigma_\beta,k_i).$$

It follows that

$$\begin{aligned}&\frac{\partial E_{\mathbf{a}}[W_{12}(k)]}{\partial e_{1\alpha}^{1}}=\\ &E_{\mathbf{a}}\left[p_{\alpha\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}_{1\alpha}^{E}(\alpha,\beta,k)}{\partial e_{1\alpha}^{1}}+d_{2\beta}\frac{\partial\tilde{a}_{1\beta}^{E}(\alpha,\beta,k)}{\partial e_{1\alpha}^{1}}\right)\right]\\ &+E_{\mathbf{a}}\left[p_{\alpha\neg\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}_{1\alpha}^{E}(\alpha,\neg\beta,k)}{\partial e_{1\alpha}^{1}}+d_{2\beta}\frac{\partial\tilde{a}_{1\beta}^{E}(\alpha,\neg\beta,k)}{\partial e_{1\alpha}^{1}}\right)\right]\\ &+E_{\mathbf{a}}\left[p_{\neg\alpha\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}_{1\alpha}^{E}(\neg\alpha,\beta,k)}{\partial e_{1\alpha}^{1}}+d_{2\beta}\frac{\partial\tilde{a}_{1\beta}^{E}(\neg\alpha,\beta,k)}{\partial e_{1\alpha}^{1}}\right)\right]\\ &+E_{\mathbf{a}}\left[p_{\neg\alpha\neg\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}_{1\alpha}^{E}(\neg\alpha,\neg\beta,k)}{\partial e_{1\alpha}^{1}}+d_{2\beta}\frac{\partial\tilde{a}_{1\beta}^{E}(\neg\alpha,\neg\beta,k)}{\partial e_{1\alpha}^{1}}\right)\right]\end{aligned}$$

$$
\begin{aligned}
&= E_{\mathbf{a}}\left[p_{\alpha\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}^{E}_{2\alpha}(\alpha,\beta,k)}{\partial e^{1}_{2\alpha}}+d_{2\beta}\frac{\partial\tilde{a}^{E}_{2\beta}(\alpha,\beta,k)}{\partial e^{1}_{2\alpha}}\right)\right]\\
&\quad +E_{\mathbf{a}}\left[p_{\alpha\neg\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}^{E}_{2\alpha}(\alpha,\neg\beta,k)}{\partial e^{1}_{2\alpha}}+d_{2\beta}\frac{\partial\tilde{a}^{E}_{2\beta}(\alpha,\neg\beta,k)}{\partial e^{1}_{2\alpha}}\right)\right]\\
&\quad +E_{\mathbf{a}}\left[p_{\neg\alpha\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}^{E}_{2\alpha}(\neg\alpha,\beta,k)}{\partial e^{1}_{2\alpha}}+d_{2\beta}\frac{\partial\tilde{a}^{E}_{2\beta}(\neg\alpha,\beta,k)}{\partial e^{1}_{2\alpha}}\right)\right]\\
&\quad +E_{\mathbf{a}}\left[p_{\neg\alpha\neg\beta}\left(d_{2\alpha}\frac{\partial\tilde{a}^{E}_{2\alpha}(\neg\alpha,\neg\beta,k)}{\partial e^{1}_{2\alpha}}+d_{2\beta}\frac{\partial\tilde{a}^{E}_{2\beta}(\neg\alpha,\neg\beta,k)}{\partial e^{1}_{2\alpha}}\right)\right]\\
&= \frac{\partial E_{\mathbf{a}}[W_{22}(k)]}{\partial e^{1}_{2\alpha}}.
\end{aligned}
$$

Analogously, one can show that $\frac{\partial E_{\mathbf{a}}[W_{12}(k)]}{\partial e^{1}_{1\beta}} = \frac{\partial E_{\mathbf{a}}[W_{22}(k)]}{\partial e^{1}_{2\beta}}$. It follows that the equilibrium is symmetric, i.e., $e^{1}_{1\alpha} = e^{1}_{2\alpha} =: e^{1}_{\alpha}$ and $e^{1}_{1\beta} = e^{1}_{2\beta} =: e^{1}_{\beta}$. Given the above expressions for $\tilde{a}^{E}_{i\alpha}(\sigma_\alpha,\sigma_\beta,k_i)$ and $\tilde{a}^{E}_{i\beta}(\sigma_\alpha,\sigma_\beta,k_i)$, the first-order condition characterizing e^{1}_{α} can be stated as

$$
\begin{aligned}
&\kappa'\left(e^{1}_{\alpha}\right) = G'(0)\,\frac{q\,(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\left(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)]\right)+p_{\alpha}\frac{d_{2\alpha}}{d_{1\alpha}}q\\
&+0.5p_{\alpha\neg\beta}d_{2\beta}\left(\frac{\partial}{\partial e^{1}_{i\alpha}}E_{\mathbf{a}}\left[a_{i\beta}\middle|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega\left(\frac{y^{\alpha}_{i11}-y^{\alpha}_{l11}-q\left(\tilde{e}^{1}_{i\alpha}-\tilde{e}^{1}_{l\alpha}\right)}{d_{1\alpha}}\right)\right]\right]\right)\\
&+0.5p_{\alpha\neg\beta}d_{2\beta}\left(\frac{\partial}{\partial e^{1}_{i\alpha}}E_{\mathbf{a}}\left[a_{i\beta}\middle|\,a_{i\beta}\in\left(a_{l\beta}+\Omega\left(\frac{y^{\alpha}_{i11}-y^{\alpha}_{l11}-q\left(\tilde{e}^{1}_{i\alpha}-\tilde{e}^{1}_{l\alpha}\right)}{d_{1\alpha}}\right),\bar{a}\right]\right]\right).
\end{aligned}
$$

The sum of the two derivatives of the conditional expectations in the second and third lines can be restated as

$$
\begin{aligned}
&\frac{\partial}{\partial e^{1}_{i\alpha}}E_{\mathbf{a}}\left[a_{i\beta}\middle|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right)\right]\right]\\
&+\frac{\partial}{\partial e^{1}_{i\alpha}}E_{\mathbf{a}}\left[a_{i\beta}\middle|\,a_{i\beta}\in\left(a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right),\bar{a}\right]\right]\\
&= \frac{\partial}{\partial e^{1}_{i\alpha}}\frac{\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right)}a_{i\beta}f(a_{i\beta})\,da_{i\beta}f(a_{l\beta})\,da_{l\beta}f(a_{i\alpha})\,da_{i\alpha}f(a_{l\alpha})\,da_{l\alpha}}{\Pr\left\{a_{i\beta}\le a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right)\right\}}\\
&+\frac{\partial}{\partial e^{1}_{i\alpha}}\frac{\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right)}^{\bar{a}}a_{i\beta}f(a_{i\beta})\,da_{i\beta}f(a_{l\beta})\,da_{l\beta}f(a_{i\alpha})\,da_{i\alpha}f(a_{l\alpha})\,da_{l\alpha}}{\Pr\left\{a_{i\beta}> a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^{1}_{i\alpha}-e^{1}_{l\alpha}-\tilde{e}^{1}_{i\alpha}+\tilde{e}^{1}_{l\alpha}}{d_{1\alpha}}\right)\right\}}
\end{aligned}
$$

$$= \frac{\partial}{\partial e^1_{i\alpha}} \frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{\Pr\left\{(d_{1\beta}-d_{2\beta})(a_{i\beta}-a_{l\beta})+(d_{2\alpha}-d_{1\alpha})(a_{l\alpha}-a_{i\alpha}) \leq (d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right\}}$$

$$+\frac{\partial}{\partial e^1_{i\alpha}} \frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)}^{\bar{a}} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{\Pr\left\{(d_{1\beta}-d_{2\beta})(a_{i\beta}-a_{l\beta})+(d_{2\alpha}-d_{1\alpha})(a_{l\alpha}-a_{i\alpha}) > (d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right\}}$$

$$= \frac{\partial}{\partial e^1_{i\alpha}} \frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)}$$

$$+\frac{\partial}{\partial e^1_{i\alpha}} \frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{a_{l\beta}+\Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)}^{\bar{a}} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{1-G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)},$$

where $\Pr\{A\}$ denotes the probability of an event A. Define $\Upsilon := \Omega\left(a_{i\alpha}-a_{l\alpha}+q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)$. The expression can be further restated as

$$\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} (a_{l\beta}+\Upsilon) f(a_{l\beta}+\Upsilon) f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)} \frac{\Omega q}{d_{1\alpha}}$$
$$-\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{a_{l\beta}+\Upsilon} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{\left(G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)\right)^2} \cdot$$
$$G'\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$
$$-\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} (a_{l\beta}+\Upsilon) f(a_{l\beta}+\Upsilon) f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{1-G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)} \frac{\Omega q}{d_{1\alpha}}$$
$$+\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{a_{l\beta}+\Upsilon}^{\bar{a}} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{\left(1-G\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)\right)^2} \cdot$$
$$G'\left((d_{2\alpha}-d_{1\alpha})\left(q\frac{e^1_{i\alpha}-e^1_{l\alpha}-\tilde{e}^1_{i\alpha}+\tilde{e}^1_{l\alpha}}{d_{1\alpha}}\right)\right)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}.$$

In equilibrium, $e^1_{ij} = \tilde{e}^1_{ij}$ for all i and j, so that $\Upsilon := \Omega(a_{i\alpha}-a_{l\alpha})$ and the expression becomes

$$= \frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} (a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})) f(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})) f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{G(0)} \frac{\Omega q}{d_{1\alpha}}$$
$$-\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{(G(0))^2} G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$
$$-\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} (a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})) f(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})) f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{1-G(0)} \frac{\Omega q}{d_{1\alpha}}$$
$$+\frac{\int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^{\bar{a}} \int_{a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})}^{\bar{a}} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{(1-G(0))^2} G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$

$$= -2\frac{\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{G(0)} G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$
$$+2\frac{\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{\underline{a}}^{\bar{a}}\int_{a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})}^{\bar{a}} a_{i\beta} f(a_{i\beta})\, da_{i\beta} f(a_{l\beta})\, da_{l\beta} f(a_{i\alpha})\, da_{i\alpha} f(a_{l\alpha})\, da_{l\alpha}}{1-G(0)} G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$

$$= -2E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in [\underline{a}, a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})]]\, G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$
$$+2E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in (a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}), \bar{a}]]\, G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}$$
$$= 2G'(0)\frac{(d_{2\alpha}-d_{1\alpha})q}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in (a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}), \bar{a}]] - E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in [\underline{a}, a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})]]\right).$$

To sum up, the first-order condition characterizing e_α^1 becomes

$$\kappa'\left(e_\alpha^1\right) = G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\left(E_{\mathbf{a}}[W_{12}(2)] - E_{\mathbf{a}}[W_{12}(1)]\right) + p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q$$
$$+p_{\alpha\neg\beta}G'(0)\frac{qd_{2\beta}(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\cdot$$
$$\left(E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in (a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}), \bar{a}]] - E_{\mathbf{a}}\left[a_{i\beta}\right| a_{i\beta} \in [\underline{a}, a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})]]\right).$$

Analogously, one can show that e_β^1 is characterized by the first-order condition

$$\kappa'\left(e_\beta^1\right) = -G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\left(E_{\mathbf{a}}[W_{12}(2)] - E_{\mathbf{a}}[W_{12}(1)]\right) + p_\beta\frac{d_{2\beta}}{d_{1\beta}}q$$
$$+p_{\neg\alpha\beta}G'(0)\frac{qd_{2\alpha}(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\cdot$$
$$\left(E_{\mathbf{a}}\left[a_{i\alpha}\right| a_{i\alpha} \in \left(a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}), \bar{a}\right]\right] - E_{\mathbf{a}}\left[a_{i\alpha}\right| a_{i\alpha} \in \left[\underline{a}, a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right]\left]\right).$$

v) In part i) of this proof, it is shown that period-2 wages are of the form

$$w_{i2}(\sigma_\alpha, \sigma_\beta, k_i) = c_2 + d_{2\alpha}\tilde{a}_{i\alpha}^E(\sigma_\alpha, \sigma_\beta, k_i) + d_{2\beta}\tilde{a}_{i\beta}^E(\sigma_\alpha, \sigma_\beta, k_i).$$

The proof continues by rewriting $\tilde{a}_{i\alpha}^E(\sigma_\alpha, \sigma_\beta, k_i)$ and $\tilde{a}_{i\beta}^E(\sigma_\alpha, \sigma_\beta, k_i)$ in all possible situations. Given that all firms correctly anticipate both workers' efforts (i.e., $\tilde{e}_{ij}^1 = e_{ij}^1$), the expression derived in part iv) can be further simplified. The following formulas are obtained:

$$\begin{aligned}
&\tilde{a}^{E}_{i\alpha}(\alpha,\beta,1)=\tilde{a}^{E}_{i\alpha}(\alpha,\beta,2)=\tilde{a}^{E}_{i\alpha}(\alpha,\neg\beta,1)=\tilde{a}^{E}_{i\alpha}(\alpha,\neg\beta,2)=a_{i\alpha},\\
&\tilde{a}^{E}_{i\beta}(\alpha,\beta,1)=\tilde{a}^{E}_{i\beta}(\alpha,\beta,2)=\tilde{a}^{E}_{i\beta}(\neg\alpha,\beta,1)=\tilde{a}^{E}_{i\beta}(\neg\alpha,\beta,2)=a_{i\beta},\\
&\tilde{a}^{E}_{i\alpha}(\neg\alpha,\beta,2)=E_{a_{1\alpha},a_{2\alpha}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right],\\
&\tilde{a}^{E}_{i\alpha}(\neg\alpha,\beta,1)=E_{a_{1\alpha},a_{2\alpha}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right)\right],\\
&\tilde{a}^{E}_{i\alpha}(\neg\alpha,\neg\beta,2)=E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right],\\
&\tilde{a}^{E}_{i\alpha}(\neg\alpha,\neg\beta,1)=E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right)\right],\\
&\tilde{a}^{E}_{i\beta}(\alpha,\neg\beta,2)=E_{a_{1\beta},a_{2\beta}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})\right]\right],\\
&\tilde{a}^{E}_{i\beta}(\alpha,\neg\beta,1)=E_{a_{1\beta},a_{2\beta}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}\right]\right],\\
&\tilde{a}^{E}_{i\beta}(\neg\alpha,\neg\beta,2)=E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})\right]\right],\text{ and}\\
&\tilde{a}^{E}_{i\beta}(\neg\alpha,\neg\beta,1)=E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}\right]\right].
\end{aligned}$$

■

Proof of Proposition 3. Begin with part a). If $p_\alpha = p_\beta = 1$, observe that $p_{\alpha\beta} = 1$ and $p_{\alpha\neg\beta} = p_{\neg\alpha\beta} = p_{\neg\alpha\neg\beta} = 0$. Proposition 2 then implies that $E_{\mathbf{a}}[W_{12}(2)] = E_{\mathbf{a}}[W_{12}(1)]$, and efforts are characterized by

$$q\frac{d_{2\alpha}}{d_{1\alpha}}=\kappa'\left(e^1_\alpha\right)\text{ and }q\frac{d_{2\beta}}{d_{1\beta}}=\kappa'\left(e^1_\beta\right).$$

Since $q\frac{d_{2\alpha}}{d_{1\alpha}} > q > q\frac{d_{2\beta}}{d_{1\beta}} > 0$, it follows that $e^1_\alpha > e^{fb} > e^1_\beta > 0$.

To prove part b), suppose that $\min\{p_\alpha, p_\beta\} < 1$ and that workers choose $\left(e^1_\alpha, e^1_\beta\right)$ with $e^1_\alpha < 0$. Such a choice requires that

$$\begin{gathered}
G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\left(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)]\right)+p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q+p_{\alpha\neg\beta}G'(0)\frac{qd_{2\beta}(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\cdot\\
\left(E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})\right]\right]\right)<0.
\end{gathered}$$

Because $\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}} > 0$ and the second and third summand on the left-hand-side of the inequality are nonnegative, this can only be the case if

$$G'(0)\left(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)]\right)<0.$$

Then, however,

$$-G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\left(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)]\right)>0$$

implying

$$-G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}(2)\right]-E_{\mathbf{a}}\left[W_{12}(1)\right]\right)+p_{\beta}\frac{d_{2\beta}}{d_{1\beta}}q+p_{\neg\alpha\beta}G'(0)\frac{qd_{2\alpha}(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\cdot$$
$$\left(E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left(a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right]\right]\right)>0\Rightarrow e^1_{\beta}>0.$$

Hence, it is impossible to observe $e^1_\alpha<0$ and $e^1_\beta<0$ simultaneously.

To see that $e^1_\alpha<0$ or $e^1_\beta<0$ is possible, suppose that $p_\alpha=p_\beta=0$, in which case $p_{\alpha\beta}=p_{\alpha\neg\beta}=p_{\neg\alpha\beta}=0$ and $p_{\neg\alpha\neg\beta}=1$. The first-order conditions characterizing optimal efforts then simplify to

$$G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,2)\right]-E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,1)\right]\right)=\kappa'\left(e^1_\alpha\right)$$

and

$$-G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\left(E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,2)\right]-E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,1)\right]\right)=\kappa'\left(e^1_\beta\right),$$

with

$$\begin{aligned}&E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,2)\right]-E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,1)\right]=\\&d_{2\alpha}E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right]+d_{2\beta}E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})\right]\right]\\&-d_{2\alpha}E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right)\right]-d_{2\beta}E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}\right]\right].\end{aligned}$$

Workers therefore choose $e^1_\alpha<0$ if $E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,2)\right]<E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,1)\right]$. If $E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,2)\right]>E_{\mathbf{a}}\left[w_{12}(\neg\alpha,\neg\beta,1)\right]$, they choose $e^1_\beta<0$. ■

Proof of Proposition 4. Starting with parts a) and b), define

$$\begin{aligned}RHS_\alpha\;:=\;&G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\left(E_{\mathbf{a}}\left[W_{12}(2)\right]-E_{\mathbf{a}}\left[W_{12}(1)\right]\right)+p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q\\&+p_{\alpha\neg\beta}G'(0)\frac{qd_{2\beta}(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\cdot\\&\left(E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\beta}|\,a_{i\beta}\in\left[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})\right]\right]\right)\end{aligned}$$

and

$$\begin{aligned}RHS_\beta\;:=\;&-G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\left(E_{\mathbf{a}}\left[W_{12}(2)\right]-E_{\mathbf{a}}\left[W_{12}(1)\right]\right)+p_\beta\frac{d_{2\beta}}{d_{1\beta}}q\\&+p_{\neg\alpha\beta}G'(0)\frac{qd_{2\alpha}(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\cdot\\&\left(E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left(a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right]\right]\right).\end{aligned}$$

The following is obtained:

$$
\begin{aligned}
\frac{\partial RHS_\alpha}{\partial p_\alpha} &= G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\frac{\partial(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)])}{\partial p_\alpha}+q\frac{d_{2\alpha}}{d_{1\alpha}} \\
&+(1-p_\beta)G'(0)\frac{qd_{2\beta}(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\cdot \\
&(E_{\mathbf{a}}[a_{i\beta}|\,a_{i\beta}\in(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}]]-E_{\mathbf{a}}[a_{i\beta}|\,a_{i\beta}\in[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})]]), \\
\frac{\partial RHS_\alpha}{\partial p_\beta} &= G'(0)\frac{q(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\frac{\partial(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)])}{\partial p_\beta} \\
&-p_\alpha G'(0)\frac{qd_{2\beta}(d_{2\alpha}-d_{1\alpha})}{d_{1\alpha}}\cdot \\
&(E_{\mathbf{a}}[a_{i\beta}|\,a_{i\beta}\in(a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha}),\bar{a}]]-E_{\mathbf{a}}[a_{i\beta}|\,a_{i\beta}\in[\underline{a},a_{l\beta}+\Omega(a_{i\alpha}-a_{l\alpha})]]), \\
\frac{\partial RHS_\beta}{\partial p_\alpha} &= -G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\frac{\partial(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)])}{\partial p_\alpha} \\
&-p_\beta G'(0)\frac{qd_{2\alpha}(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\cdot \\
&\left(E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left(a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right]\right]\right), \\
\frac{\partial RHS_\beta}{\partial p_\beta} &= -G'(0)\frac{q(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\frac{\partial(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)])}{\partial p_\beta}+q\frac{d_{2\beta}}{d_{1\beta}} \\
&+(1-p_\alpha)G'(0)\frac{qd_{2\alpha}(d_{1\beta}-d_{2\beta})}{d_{1\beta}}\cdot \\
&\left(E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left(a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{i\alpha}|\,a_{i\alpha}\in\left[\underline{a},a_{l\alpha}+\frac{1}{\Omega}(a_{i\beta}-a_{l\beta})\right]\right]\right).
\end{aligned}
$$

It can further be shown that

$$
\begin{aligned}
&\frac{\partial(E_{\mathbf{a}}[W_{12}(2)]-E_{\mathbf{a}}[W_{12}(1)])}{\partial p_\alpha} \\
=\;& E_{\mathbf{a}}[p_\beta w_{12}(\alpha,\beta,2)+(1-p_\beta)w_{12}(\alpha,\neg\beta,2)-p_\beta w_{12}(\neg\alpha,\beta,2)-(1-p_\beta)w_{12}(\neg\alpha,\neg\beta,2)] \\
&+E_{\mathbf{a}}[-p_\beta w_{12}(\alpha,\beta,1)-(1-p_\beta)w_{12}(\alpha,\neg\beta,1)+p_\beta w_{12}(\neg\alpha,\beta,1)+(1-p_\beta)w_{12}(\neg\alpha,\neg\beta,1)] \\
=\;& (1-p_\beta)E_{\mathbf{a}}[(w_{12}(\alpha,\neg\beta,2)-w_{12}(\alpha,\neg\beta,1))-(w_{12}(\neg\alpha,\neg\beta,2)-w_{12}(\neg\alpha,\neg\beta,1))] \\
&-p_\beta E_{\mathbf{a}}[(w_{12}(\neg\alpha,\beta,2)-w_{12}(\neg\alpha,\beta,1))].
\end{aligned}
$$

Inserting for $w_{12}(\sigma_\alpha,\sigma_\beta,k_i)$ the expressions determined in Proposition 2, the expression in the lower lines can be restated as

$$
\begin{aligned}
&(1-p_\beta)d_{2\beta}(E_{\mathbf{a}}[a_{1\beta}|\,a_{1\beta}\in[\underline{a},a_{2\beta}+\Omega(a_{1\alpha}-a_{2\alpha})]]-E_{\mathbf{a}}[a_{1\beta}|\,a_{1\beta}\in(a_{2\beta}+\Omega(a_{1\alpha}-a_{2\alpha}),\bar{a}]]) \\
&-(1-p_\beta)d_{2\alpha}\left(E_{\mathbf{a}}\left[a_{1\alpha}|\,a_{1\alpha}\in\left[a_{2\alpha}+\frac{1}{\Omega}(a_{1\beta}-a_{2\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{1\alpha}|\,a_{1\alpha}\in\left[\underline{a},a_{2\alpha}+\frac{1}{\Omega}(a_{1\beta}-a_{2\beta})\right)\right]\right) \\
&-(1-p_\beta)d_{2\beta}(E_{\mathbf{a}}[a_{1\beta}|\,a_{1\beta}\in[\underline{a},a_{2\beta}+\Omega(a_{1\alpha}-a_{2\alpha})]]-E_{\mathbf{a}}[a_{1\beta}|\,a_{1\beta}\in(a_{2\beta}+\Omega(a_{1\alpha}-a_{2\alpha}),\bar{a}]]) \\
&-p_\beta d_{2\alpha}\left(E_{\mathbf{a}}\left[a_{1\alpha}|\,a_{1\alpha}\in\left[a_{2\alpha}+\frac{1}{\Omega}(a_{1\beta}-a_{2\beta}),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{1\alpha}|\,a_{1\alpha}\in\left[\underline{a},a_{2\alpha}+\frac{1}{\Omega}(a_{1\beta}-a_{2\beta})\right)\right]\right)
\end{aligned}
$$

$$= -d_{2\alpha}\left(E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}\in\left[a_{2\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}\in\left[\underline{a},a_{2\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right)\right]\right)<0.$$

Because all the a_{ij} are iid, it is further observed that

$$\begin{aligned}
& E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}\in\left[a_{2\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}\in\left[\underline{a},a_{2\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right)\right] \\
= \;& E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}\in\left[a_{2\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right),\bar{a}\right]\right]-E_{\mathbf{a}}\left[a_{2\alpha}|\, a_{2\alpha}\in\left[\underline{a},a_{1\alpha}+\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right)\right] \\
= \;& E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]-E_{\mathbf{a}}\left[a_{2\alpha}|\, a_{2\alpha}-a_{1\alpha}\leq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right] \\
= \;& E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]-E_{\mathbf{a}}\left[a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{2\beta}-a_{1\beta}\right)\right] \\
= \;& E_{\mathbf{a}}\left[a_{1\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]-E_{\mathbf{a}}\left[a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right] \\
= \;& E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right].
\end{aligned}$$

The other differences between the conditional expectations can be transformed in a similar way. Hence,

$$\begin{aligned}
\frac{\partial e_{\alpha}^{1}}{\partial p_{\alpha}} &> 0\Leftrightarrow\frac{d_{2\alpha}}{d_{1\alpha}}+\left(1-p_{\beta}\right)G'\left(0\right)\frac{d_{2\beta}\left(d_{2\alpha}-d_{1\alpha}\right)}{d_{1\alpha}}E_{\mathbf{a}}\left[a_{1\beta}-a_{2\beta}|\, a_{1\beta}-a_{2\beta}\geq\Omega\left(a_{1\alpha}-a_{2\alpha}\right)\right] \\
&> G'\left(0\right)\left(d_{2\alpha}-d_{1\alpha}\right)\frac{d_{2\alpha}}{d_{1\alpha}}E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial e_{\beta}^{1}}{\partial p_{\alpha}} &> 0\Leftrightarrow G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}d_{2\alpha}E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right] \\
&-p_{\beta}G'\left(0\right)\frac{qd_{2\alpha}\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]>0 \\
&\Leftrightarrow G'\left(0\right)\frac{q\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}d_{2\alpha}E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right]\left(1-p_{\beta}\right)>0.
\end{aligned}$$

It follows that $\frac{\partial e_{\beta}^{1}}{\partial p_{\alpha}}\geq 0$, with a strict inequality for $p_{\beta}<1$. Analogously, it can be shown that

$$\frac{\partial\left(E_{\mathbf{a}}\left[W_{12}\left(2\right)\right]-E_{\mathbf{a}}\left[W_{12}\left(1\right)\right]\right)}{\partial p_{\beta}}=d_{2\beta}E_{\mathbf{a}}\left[a_{1\beta}-a_{2\beta}|\, a_{1\beta}-a_{2\beta}\geq\Omega\left(a_{1\alpha}-a_{2\alpha}\right)\right]>0,$$

implying $\frac{\partial e_{\alpha}^{1}}{\partial p_{\beta}}\geq 0$, with a strict inequality for $p_{\alpha}<1$, and

$$\begin{aligned}
\frac{\partial e_{\beta}^{1}}{\partial p_{\beta}} &> 0\Leftrightarrow\frac{d_{2\beta}}{d_{1\beta}}+\left(1-p_{\alpha}\right)G'\left(0\right)\frac{d_{2\alpha}\left(d_{1\beta}-d_{2\beta}\right)}{d_{1\beta}}E_{\mathbf{a}}\left[a_{1\alpha}-a_{2\alpha}|\, a_{1\alpha}-a_{2\alpha}\geq\frac{1}{\Omega}\left(a_{1\beta}-a_{2\beta}\right)\right] \\
&> G'\left(0\right)\left(d_{1\beta}-d_{2\beta}\right)\frac{d_{2\beta}}{d_{1\beta}}E_{\mathbf{a}}\left[a_{1\beta}-a_{2\beta}|\, a_{1\beta}-a_{2\beta}\geq\Omega\left(a_{1\alpha}-a_{2\alpha}\right)\right].
\end{aligned}$$

Turning to part c), suppose that $p_\alpha = p_\beta = 1$. In this case the efforts e^1_α and e^1_β are characterized by

$$q\frac{d_{2\alpha}}{d_{1\alpha}} = \kappa'\left(e^1_\alpha\right)$$

and

$$q\frac{d_{2\beta}}{d_{1\beta}} = \kappa'\left(e^1_\beta\right),$$

which means that $e^1_\alpha > e^{fb} > e^1_\beta$. Further suppose that

$$1 > G'(0)\left(d_{2\alpha} - d_{1\alpha}\right) E_{\mathbf{a}}\left[a_{1\alpha} - a_{2\alpha} \middle| a_{1\alpha} - a_{2\alpha} \geq \frac{1}{\Omega}\left(a_{1\beta} - a_{2\beta}\right)\right],$$

which implies $\frac{\partial e^1_\alpha}{\partial p_\alpha} > 0$ for $p_\beta = 1$. If, starting from $p_\alpha = p_\beta = 1$, the probability p_α is reduced, then e^1_α becomes smaller, whereas e^1_β does not change. For a sufficiently small reduction in p_α, e^1_α gets closer to the efficient level, leading to an increase in welfare. With a similar argumentation, if

$$1 < G'(0)\left(d_{1\beta} - d_{2\beta}\right) E_{\mathbf{a}}\left[a_{1\beta} - a_{2\beta} \middle| a_{1\beta} - a_{2\beta} \geq \Omega\left(a_{1\alpha} - a_{2\alpha}\right)\right]$$

a sufficiently small decrease in p_β leads to a more efficient effort e^1_β, while having no effect on e^1_α. ■

Proof of Proposition 5. The proof is analogous to that of Proposition 2. The $\tilde{a}^E_{i\alpha}\left(\sigma_\alpha, k_i\right)$'s are given by

$$\begin{aligned}
\tilde{a}^E_{i\alpha}(\alpha, 1) &= \tilde{a}^E_{i\alpha}(\alpha, 2) = a_{i\alpha}, \\
\tilde{a}^E_{i\alpha}(\neg\alpha, 2) &= E_{\hat{\mathbf{a}}}\left[a_{i\alpha} \middle| a_{i\alpha} \in \left[a_{l\alpha}, \bar{a}\right]\right], \\
\tilde{a}^E_{i\alpha}(\neg\alpha, 1) &= E_{\hat{\mathbf{a}}}\left[a_{i\alpha} \middle| a_{i\alpha} \in \left[\underline{a}, a_{l\alpha}\right)\right].
\end{aligned}$$

And with the same argumentation as in subsection 4.1, a worker's period-1 wage is given by

$$\begin{aligned}
w_1 &= c_1 + d_{1\alpha}E[a] + qe^1_\alpha - 0.5\left(E_{\hat{\mathbf{a}}}\left[W_{12}(2)\right] + E_{\hat{\mathbf{a}}}\left[W_{12}(1)\right]\right) \\
&\quad + 0.5(1+S)\left(c_2 + E_{\hat{\mathbf{a}}}\left[d_{2\alpha}a_{1\alpha} \middle| a_{1\alpha} \geq a_{2\alpha}\right]\right) + 0.5(1+S)\left(c_1 + E_{\hat{\mathbf{a}}}\left[d_{1\alpha}a_{1\alpha} \middle| a_{1\alpha} \leq a_{2\alpha}\right]\right).
\end{aligned}$$

■

Proof of Corollary 1. Since effort is determined by the condition

$$\hat{G}'(0)\frac{q\left(d_{2\alpha} - d_{1\alpha}\right)}{d_{1\alpha}}\left(E_{\hat{\mathbf{a}}}\left[W_{12}(2)\right] - E_{\hat{\mathbf{a}}}\left[W_{12}(1)\right]\right) + p_\alpha\frac{d_{2\alpha}}{d_{1\alpha}}q = \kappa'\left(e^1_\alpha\right),$$

the proof would be complete if it could be shown that $E_{\hat{\mathbf{a}}}[W_{12}(2)] \geq E_{\hat{\mathbf{a}}}[W_{12}(1)]$. Note that

$$\begin{aligned} E_{\hat{\mathbf{a}}}[W_{12}(2)] &= c_2 + d_{2\alpha} E_{\hat{\mathbf{a}}}\left[p_\alpha \tilde{a}_{1\alpha}^E(\alpha, 2) + (1 - p_\alpha)\tilde{a}_{1\alpha}^E(\neg\alpha, 2)\right] \\ &= c_2 + d_{2\alpha} E_{\hat{\mathbf{a}}}\left[p_\alpha a_{1\alpha} + (1 - p_\alpha) E_{\hat{\mathbf{a}}}\left[a_{1\alpha} \middle| a_{1\alpha} \in [a_{2\alpha}, \bar{a}]\right]\right] \end{aligned}$$

and

$$\begin{aligned} E_{\hat{\mathbf{a}}}[W_{12}(1)] &= c_2 + d_{2\alpha} E_{\hat{\mathbf{a}}}\left[p_\alpha \tilde{a}_{1\alpha}^E(\alpha, 1) + (1 - p_\alpha)\tilde{a}_{1\alpha}^E(\neg\alpha, 1)\right] \\ &= c_2 + d_{2\alpha} E_{\hat{\mathbf{a}}}\left[p_\alpha a_{1\alpha} + (1 - p_\alpha) E_{\hat{\mathbf{a}}}\left[a_{1\alpha} \middle| a_{1\alpha} \in [\underline{a}, a_{2\alpha})\right]\right]. \end{aligned}$$

Because $E_{\hat{\mathbf{a}}}[a_{1\alpha}| a_{1\alpha} \in [a_{2\alpha}, \bar{a}]] \geq E_{\hat{\mathbf{a}}}[a_{1\alpha}| a_{1\alpha} \in [\underline{a}, a_{2\alpha})]$, it is straightforward to see that $E_{\hat{\mathbf{a}}}[W_{12}(2)] \geq E_{\hat{\mathbf{a}}}[W_{12}(1)]$. ■

Proof of Proposition 6. The proof is completely analogous to that of Proposition 4 and is therefore omitted. ■

Proof of Corollary 2. a) Suppose that $p_\alpha, p_\beta > 0$. Wages conditional on the job-level are not constant, because performance on the two tasks is observed with positive probability, in which case wages depend on the inferred (and thus actual) ability levels. Suppose that $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0$, implying that worker 1 is promoted. Furthermore, suppose that performance on both tasks is observable. If $a_{1\alpha} > a_{2\alpha}$, $a_{1\beta} > a_{2\beta}$, the external firms infer these inequalities from the performance observations and, thus, offer a higher wage to the promoted worker 1. Instead, if $a_{1\alpha} < a_{2\alpha}$, $a_{1\beta} < a_{2\beta}$, the external firms correctly infer that worker 2 has higher ability on both of the tasks and offer a higher wage to the non-promoted worker 2. Notice that it is always possible to find parameter values such that $a_{1\alpha} > a_{2\alpha}$, $a_{1\beta} > a_{2\beta}$, and $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0$ or $a_{1\alpha} < a_{2\alpha}$, $a_{1\beta} < a_{2\beta}$, and $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0$ are jointly satisfied. The argumentation is analogous in the case $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) < 0$.

b) Without loss of generality, suppose $p_\alpha = 0$ and $p_\beta > 0$ and that worker 1 is promoted to job 2 (so that $(d_{2\alpha} - d_{1\alpha})(a_{1\alpha} - a_{2\alpha}) + (d_{1\beta} - d_{2\beta})(a_{2\beta} - a_{1\beta}) \geq 0$). If external firms do not observe performance on task β, period-2 wages for the two workers are given by

$$\begin{aligned} & w_{12}(\neg\alpha, \neg\beta, 2) \\ = \;& c_2 + d_{2\alpha} E_{\mathbf{a}}\left[a_{1\alpha} \middle| a_{1\alpha} \in \left[a_{2\alpha} + \frac{1}{\Omega}(a_{1\beta} - a_{2\beta}), \bar{a}\right]\right] + d_{2\beta} E_{\mathbf{a}}\left[a_{1\beta} \middle| a_{1\beta} \in [\underline{a}, a_{2\beta} + \Omega(a_{1\alpha} - a_{2\alpha})]\right] \end{aligned}$$

and

$$\begin{aligned} & w_{22}(\neg\alpha,\neg\beta,1) \\ = \; & c_2 + d_{2\alpha} E_{\mathbf{a}} \left[a_{2\alpha} \middle| \, a_{2\alpha} \in \left[\underline{a}, a_{1\alpha} + \frac{1}{\Omega}(a_{2\beta} - a_{1\beta}) \right) \right] + d_{2\beta} E_{\mathbf{a}} \left[a_{2\beta} | \, a_{2\beta} \in (a_{1\beta} + \Omega(a_{2\alpha} - a_{1\alpha}), \bar{a}] \right]. \end{aligned}$$

Instead, if external firms observe performance on task β, period-2 wages for the two workers are given by

$$w_{12}(\neg\alpha,\beta,2) = c_2 + d_{2\alpha} E_{a_{1\alpha},a_{2\alpha}} \left[a_{1\alpha} \middle| \, a_{1\alpha} \in \left[a_{2\alpha} + \frac{1}{\Omega}(a_{1\beta} - a_{2\beta}), \bar{a} \right] \right] + d_{2\beta} a_{1\beta}$$

and

$$w_{22}(\neg\alpha,\beta,1) = c_2 + d_{2\alpha} E_{a_{1\alpha},a_{2\alpha}} \left[a_{2\alpha} \middle| \, a_{2\alpha} \in \left[\underline{a}, a_{1\alpha} + \frac{1}{\Omega}(a_{2\beta} - a_{1\beta}) \right) \right] + d_{2\beta} a_{2\beta}.$$

Suppose that the parameter constellations are such that $w_{12}(\neg\alpha,\neg\beta,2) > w_{22}(\neg\alpha,\neg\beta,1)$ (this condition can be fulfilled, for example, for sufficiently high $d_{2\alpha}$ and sufficiently low $d_{2\beta}$). Hence, if external firms do not observe performance on task β, the promoted worker 1 receives a higher wage than the non-promoted worker 2. Now let performance on task β be observable. If parameter constellations are such that

$$w_{12}(\neg\alpha,\beta,2) \geq w_{22}(\neg\alpha,\beta,1)$$

for all $(a_{1\beta}, a_{2\beta}) \in [\underline{a}, \bar{a}] \times [\underline{a}, \bar{a}]$ (e.g., because $d_{2\alpha}$ is much higher than $d_{2\beta}$), worker 1 always receives a higher wage than worker 2, regardless of whether external firms observe workers' performance on task β and of the realized ability levels; in this situation, there is no wage overlap across job levels. If instead

$$w_{12}(\neg\alpha,\beta,2) < w_{22}(\neg\alpha,\beta,1)$$

for some $(a_{1\beta}, a_{2\beta}) \in [\underline{a}, \bar{a}] \times [\underline{a}, \bar{a}]$, worker 1 sometimes receives a lower wage than worker 2, meaning that wage distributions overlap across job levels.

c) If $p_\alpha = p_\beta = 0$, period-2 wages in the two jobs are deterministic (meaning that they do not depend on the ability realizations) and given by

$$w_{i2}(\neg\alpha,\neg\beta,1) = c_2 + d_{2\alpha} \tilde{a}^E_{i\alpha}(\neg\alpha,\neg\beta,1) + d_{2\beta} \tilde{a}^E_{i\beta}(\neg\alpha,\neg\beta,1)$$

and

$$w_{i2}(\neg\alpha,\neg\beta,2) = c_2 + d_{2\alpha} \tilde{a}^E_{i\alpha}(\neg\alpha,\neg\beta,2) + d_{2\beta} \tilde{a}^E_{i\beta}(\neg\alpha,\neg\beta,2).$$

Either $w_{i2}(\neg\alpha,\neg\beta,1) \geq w_{i2}(\neg\alpha,\neg\beta,2)$ or $w_{i2}(\neg\alpha,\neg\beta,1) < w_{i2}(\neg\alpha,\neg\beta,2)$, but it is not possible that the former inequality holds for some realizations of the random variables, whereas the latter holds for others. ■

9 References

Agrawal, A., C.R. Knoeber, and T. Tsoulouhas. 2006. "Are Outsiders Handicapped in CEO Successions?," 12 *Journal of Corporate Finance* 619-44.

Audas, R., T. Barmby, and J. Treble. 2004. "Luck, Effort, and Reward in an Organizational Hierarchy," 22 *Journal of Labor Economics* 379-96.

Baker, G., M. Gibbs, and B. Holmström. 1994. "The Internal Economics of the Firm: Evidence from Personnel Data," 109 *Quarterly Journal of Economics* 921-55.

Baker, G., M. Jensen, and K.J. Murphy. 1988. "Compensation and Incentives: Practice Versus Theory," 43 *Journal of Finance* 593-616.

Benson, A., D. Li, and K. Shue. 2019. "Promotions and the Peter Principle," *Quarterly Journal of Economics* (forthcoming).

Bernhardt, D. 1995. "Strategic promotion and compensation," 62 *Review of Economic Studies* 315-39.

Bognanno, M., Melero, E., 2016. "Promotion signals, experience, and education," 25 *Journal of Economics and Management Strategy* 111-32.

Brilon, S. 2015. "Job Assignment With Multivariate Skills and the Peter Principle," 32 *Labour Economics* 112-21.

Brown, J. 2011. "Quitters Never Win: The (Adverse) Incentive Effects of Competing with Superstars," 119 *Journal of Political Economy* 982-1013.

Brown, J. and D.B. Minor 2014. "Selecting the Best? Spillover and Shadows in Elimination Tournaments," 60 *Management Science* 3087-102.

Cassidy, H., J. DeVaro, and A. Kauhanen. 2016. "Promotion Signaling, Gender, and Turnover: New Theory and Evidence," 126 *Journal of Economic Behavior and Organization* 140-66.

Chowdhury, S.M., and O. Gürtler. 2015. "Sabotage in Contests: A Survey," 164 *Public Choice* 135-55.

Coffey, B., and M.T. Maloney. 2010. "The Thrill of Victory: Measuring the Incentive to Win," 28 *Journal of Labor Economics* 87-112.

Dato, S., Grunewald, A., Kräkel, M., and D. Müller. 2016. "Asymmetric employer information, promotions, and the wage policy of firms," 100 *Games and Economic Behavior* 273-300.

DeVaro, J. 2006a. "Strategic Promotion Tournaments and Worker Performance," 27 *Strategic Management Journal* 721-40.

DeVaro, J. 2006b. "Internal Promotion Competitions in Firms," 37 *RAND Journal of Economics* 521-41.

DeVaro, J., S. Ghosh, and C. Zoghi. 2018. "Job Characteristics and Labor Market Discrimination in Promotions," 57 *Industrial Relations* 389-434.

DeVaro, J., and O. Gürtler. 2016a. "Strategic Shirking in Promotion Tournaments," 32 *Journal of Law, Economics, and Organization* 620-51.

DeVaro, J., and O. Gürtler. 2016b. "Strategic Shirking: A Theoretical Analysis of Multitasking and Specialization," 57 *International Economic Review* 507-32.

DeVaro, J., and A. Kauhanen. 2016. "An "Opposing Responses" Test of Classic Versus Market-Based Promotion Tournaments," 34 *Journal of Labor Economics* 747-79.

DeVaro, J., and H. Morita. 2013. "Internal Promotion and External Recruitment: A Theoretical and Empirical Analysis," 31 *Journal of Labor Economics* 227-69.

DeVaro, J., and M. Waldman. 2012. "The Signaling Role of Promotions: Further Theory and Empirical Evidence," 30 *Journal of Labor Economics* 91-147.

Fan, X., and J. DeVaro. 2020. "Job Hopping and Adverse Selection in the Labor Market," 36 *Journal of Law, Economics, and Organization*.

Ghosh, S., and M. Waldman. 2010. "Standard Promotion Practices Versus Up-Or-Out Contracts," 41 *RAND Journal of Economics* 301-25.

Gibbons, R., and M. Waldman. 1999. "A Theory of Wage and Promotion Dynamics Inside Firms," 114 *Quarterly Journal of Economics* 1321-58.

Gibbons, R., and M. Waldman. 2006. "Enriching a Theory of Wage and Promotion Dynamics Inside Firms," 24 *Journal of Labor Economics* 59-107.

Gibbs, M. 1995. "Incentive Compensation in a Corporate Hierarchy," 19 *Journal of Accounting and Economics* 247-77.

Gürtler, M. and O. Gürtler. 2015. "The Optimality of Heterogeneous Tournaments," 33 *Journal of Labor Economics,* 1007-42.

Gürtler, M. and O. Gürtler. 2019. "Promotion Signaling, Discrimination, and Positive Discrimination Policies," *RAND Journal of Economics* (forthcoming).

Holmström, B. 1999. "Managerial Incentive Problems: A Dynamic Perspective," 66 *Review of Economic Studies* 169-82.

Holmström, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership and Job Design," 7 *Journal of Law, Economics, and Organization* 24-52.

Kauhanen, A., and S. Napari. 2012. "Career and Wage Dynamics: Evidence from Linked-Employer Employee Data," 36 *Research in Labor Economics* 35-76.

Ke, R., J. Li, and M. Powell. 2018. "Managing careers in organizations," 36 *Journal of Labor Economics* 197-252.

Kim, S., and E. Usui, "Employer Learning, Job Changes, and Wage Dynamics," 2014.

Lazear, E.P. 2000. "The Future of Personnel Economics," 110 *The Economic Journal* F611-F639.

Lazear, E.P., and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts," 89 *Journal of Political Economy* 841-64.

Lazear, E.P., Shaw, K., and C. Stanton 2018. "Who Gets Hired? The Importance of Competition Among Applicants," 36 *Journal of Labor Economics* 133-81.

List, J., D. Van Soest, J. Stoop and H. Zhou, "On the role of group size in tournaments: theory and evidence from lab and field experiments," NBER working paper 20008, 2014.

MacDonald, G. 1982. "A Market Equilibrium Theory of Job Assignments and Sequential Accumulation of Information," 72 *American Economic Review* 1038-55.

Mas-Colell, A., Whinston, M.D., and Green, J.R. 1995. *Microeconomic theory.* Oxford: Oxford University Press.

Milgrom, P., and S. Oster. 1987. "Job Discrimination, Market Forces, and the Invisibility Hypothesis," 102 *Quarterly Journal of Economics* 453-76.

Milgrom, P., and J. Roberts. 1988. "An Economic Approach to Influence Activities in Organizations," 94 *American Journal of Sociology* S154-79.

Okamura, K. 2011. "The Signalling Role of Promotions in Japan," Discussion Paper No. 1112. Graduate School of Economics, Kobe University.

Prasad, S. 2009. "Task Assignment and Incentives: Generalists Versus Specialists," 40 *RAND Journal of Economics* 380-403.

Prasad, S., and H. Tran. 2013. "Work Practices, Incentives for Skills, and Training," 23 *Labour Economics* 66-76.

Prendergast, Canice. 1993. "The Role of Promotion in Inducing Specific Human Capital Acquisition," 108 *Quarterly Journal of Economics* 108, 523-34.

Putt, A. 2006. *Putt's Law and the Successful Technocrat: How to Win in the Information Age*. United States: Wiley-IEEE Press.

Schönberg, U. 2007. "Testing for Asymmetric Employer Learning," 25 *Journal of Labor Economics*, 651-691.

Schöttner, A., and V. Thiele. 2010. "Promotion Tournaments and Individual Performance Pay," 19 *Journal of Economics and Management Strategy* 699-731.

Selten, Reinhard. 1975. "Reexamination of the perfectness concept for equilibrium points in extensive games," 4 *International Journal of Game Theory* 25-55.

Smeets, V., M. Waldman, and F. Warzynski. 2019. "Performance, Career Dynamics, and Span of Control," 37 *Journal of Labor Economics* 1183-213.

Tsoulouhas, T., C.R. Knoeber, and A. Agrawal. 2007. "Contests to Become CEO: Incentives, Selection and Handicaps," 30 *Economic Theory* 195-221.

Waldman, M. 1984. "Job Assignments, Signalling, and Efficiency," 15 *RAND Journal of Economics* 255-67.

Waldman, M. 2003. "Ex Ante Versus Ex Post Optimal Promotion Rules: The Case of Internal Promotion," 41 *Economic Inquiry* 27-41.

Waldman, M. 2013. "Classic Promotion Tournaments Versus Market-Based Tournaments," 31 *International Journal of Industrial Organization* 198-210.

Waldman, M. 2016. "The Dual Avenues of Labor Market Signaling," 41 *Labour Economics* 120-34.

Waldman, M., and O. Zax. 2016. "An Exploration of the Promotion Signaling Distortion," 32 *Journal of Law, Economics, and Organization* 119-49.

Waldman, M., and O. Zax. 2019. "Promotion Signaling and Human Capital Investments," *American Economic Journal: Microeconomics.*

Zábojník, J. 2012. "Promotion Tournaments in Market Equilibrium," 51 *Economic Theory* 213-40.

Zábojník, J., and D. Bernhardt. 2001. "Corporate Tournaments, Human Capital Acquisition, and the Firm Size-Wage Relation," 68 *Review of Economic Studies* 693-716.