Empathy wages?: Gratitude and gift exchange in employment relationships

James N. Baron
Yale School of Management, Yale University, United States

Abstract
Economists have argued that employers sometimes pay above-market premiums (efficiency wages) in order to attract, motivate, and/or retain valued personnel. Drawing on recent work examining reciprocity and gift exchange, this paper proposes the notion of “empathy wages,” in which the effect of the premium paid depends on the extent to which it elicits gratitude from recipients. We argue that a particular gift (monetary or otherwise) offered by an employer is likely to elicit more gratitude among “non-stars”: workers who are relatively disadvantaged and in the lower part of the performance distribution. In contrast to “stars,” “non-stars” are likely to compare the treatment they receive to the inferior opportunities or treatment they (might) have received outside of their present employment situation. Star workers, in contrast, are likely to believe that they are worth whatever they can command. The economic viability of such “empathy wages” thus depends on how much star versus non-star workers vary in gratitude, relative to how they differ in output and compensation. We explore a variety of data bearing on how much stars differ from non-stars in their respective output and earnings (in star contexts such as professional sports and real estate sales). We then review or reanalyze some prior studies on gift exchange, documenting that those who are relatively disadvantaged and/or low performers do appear more grateful (or inclined to reciprocate gifts) than stars. Indeed, the magnitude of the difference is sufficiently large that it could offset quite marked differences in productivity or quite small differences in compensation (both of which would make stars relatively more attractive to employers). We suggest some conditions under which gratitude-based employment systems are more likely to flourish in real-world settings, as well as some fruitful lines for future research on these topics.

© 2013 Elsevier Ltd. All rights reserved.

Contents
1. Introduction ................................................................. 114
2. Arguments for stars .................................................. 116
3. Gift exchange and prospects for gratitude ......................... 117
4. Relative performance and pay among star employees ............. 119
4.1. Performance differentials ......................................... 119
December 1, 2009: “Sirius chief executive Mel Karmazin offered a sneak peek this week into upcoming contract negotiations with Howard Stern...Stern’s $500 million contract with the satellite radio giant expires at the end of 2010, and it’s doubtful [he] will command the same kind of money this time around. (Sirius, which has 18.5 million subscribers, was nearly forced to declare bankruptcy earlier this year.)

“The hope would be that we would come out with Howard staying with our service,” Karmazin said, adding: “You have to now assume that the negotiations are at a stage where everything is in [the media], so if I were to say, yes, we got every penny’s worth, Howard would come in with that piece of paper and say ‘See? I sold myself too cheap.’”

February 19, 2010: “I convinced myself that normal rules didn’t apply...I felt I was entitled” (Tiger Woods).²

1. Introduction

It is a truism to observe that our culture has become preoccupied with star performers—in sports, entertainment, commerce, education, and just about everywhere else. This obsession with top performers in our popular discourse is illustrated in Fig. 1, which documents how media coverage of business over the first decade of the 21st century increasingly made reference to “stars” and “talent.”¹ Rapt attention to stars is equally prevalent in the academic literature. For instance, burgeoning interest in the economics of superstars and winner-take-all-markets is reflected in a new entry entitled “superstars, economics of” in the most recent edition of The New Palgrave Dictionary of Economics (Oi, 2008).

Indeed, interest in how to capitalize on the upper end of the talent distribution has been a focus of both practitioners and academics interested in labor markets. In 1997, the global consulting firm McKinsey famously proclaimed the existence of a “war for talent” that constituted a strategic business challenge and key driver of organizational performance, declaring unequivocally that “Talent Wins” (see Michaels, Handfield-Jones, & Axelrod, 1997). Performance-based pay regimes have diffused broadly between and within organizations, bolstered by academic studies showing that they not only boost productivity among an organization’s existing workforce, but also help attract and retain star employees (e.g., Lazear, 2000).

Another strand of labor economics has examined so-called efficiency wages as a means of attracting and retaining stellar performers. Economists have also proposed various versions of the efficiency wage hypothesis. It is generally assumed that workers vary considerably in ability, the willingness to supply effort, and/or the propensity to remain loyal to an employer. Thus, when workers can easily change jobs, it can be efficient for employers to offer premium wages as a means of attracting and retaining premium talent.

Empirical studies have offered considerable support for the hypothesis (e.g., Efficiency Wage, 2013; Katz, 1986). A spirited debate continues, however, regarding the economic, psychological, and/or sociological micro-foundations of efficiency wages. Initial variants of the hypothesis focused on the role of above-market wages in encouraging sorting by ability; reducing shirking and the costs of monitoring; lowering turnover; and, under conditions of poverty, increasing productivity by improving nutrition and health. More recent versions have tended to highlight psychological and sociological rationales for wages above market-clearing, invoking concepts such as fairness (e.g., what can the employer afford to pay), reciprocity, and gift exchange (Akerlof, 1982; Bewley, 1999; Rabin, 1993).

Yet there have been plenty of explicit and implicit critiques of this view within scholarly, business, and...
popular literatures. The potential for salient extrinsic rewards to crowd out intrinsic motivation, well documented by psychologists (Deci, 1975; Greene, Sternberg, & Lepper, 1976; Staw, Calder, Hess, & Sandelands, 1980; Tang & Hall, 1995), has received increasing attention from practitioners in business (e.g., Pink, 2011), education and child rearing (Kohn, 1993, 2005), and other spheres. The pre-eminent focus on stars has also been under assault, particularly in the post-Lehman Brothers era, with scholars and practitioners alike questioning human resource strategies based on excessive reliance on a small cadre of highly compensated superstars (Groysberg, 2012). Numerous academic studies and corporate testimonials extol the virtues of so-called high commitment work systems (Appelbaum, Bailey, Berg, Kalleberg, & 2000; Pfeffer, 1994; Beer, 2012; Mackey and Sisodia, 2013), emphasizing prospects for “raising productivity and profitability by creating better working conditions for the worst off” (Heymann, 2010, p. 1). Corporate leaders who advocate more altruistic and dignified approaches to human resources have received immense publicity in recent years, such as John Mackey of Whole Foods, Jim Sinegal of Costco, George Zimmer of Mens Wearhouse, Ben Cohen and Jerry Greenstein of Ben & Jerry’s, Jim Goodnight of SAS, and Tony Hsieh of Zappos.

In this essay, I suggest that taking the concepts of gift exchange and reciprocal altruism seriously might actually lead to predictions that are the opposite of what conventional efficiency wage theories hypothesize. The basic idea can be stated simply: the impact of efficiency wages depends on the psychological and sociological bang for the buck associated with the premium compensation or treatment being provided—the employer’s capacity to generate feelings of gratitude and obligation by engaging in what is perceived to be benevolent and empathetic treatment of employees, who in turn feel compelled to reciprocate through some aspect(s) of their job performance. I will provisionally call this the empathy wage hypothesis. I argue that prospects for eliciting gratitude may potentially be much greater in relative terms toward the bottom of the talent distribution, whereas creating equivalent feelings at the top of the distribution is much more difficult and costly. Efficiency wages paid to star performers may dissuade them from seeking to leave the firm, but they are unlikely to affirmatively bind the employee to the enterprise: like radio personality Howard Stern in the quotation above, the star is likely to perceive that he or she is worth whatever the employer was willing to pay.

At the other end of the talent distribution, if empathy wages can create powerful feelings of gratitude and attachment, this not only makes it costly for an employee to change jobs but also creates powerful inducements to want to stay on the job and perform well. An employee’s return from star versus non-star employees will thus depend not only on the relative ability and market wages of the two groups, but also on their relative responsiveness to discretionary compensation and treatment that surpasses what is perceived as typical within the labor market.

4 Note that compensation in efficiency wage theories need not be restricted to pay and benefits. Superior personal treatment, enhanced recognition, greater job latitude, and a host of other perquisites that are nonstandard in the industry and that entail cost (or opportunity cost) from the employer’s perspective could constitute other means of premium treatment.
This alternative perspective is capable of explaining the abiding presence of above-market wage regimes in contexts that are somewhat baffling from the vantage point of conventional theory. As an example, consider Men’s Wearhouse, an “everyday low price” retailer of men’s clothing (primarily business apparel). The firm has competed very successfully over the last few decades in a singularly unattractive industry setting: small margins and a decline in absolute size of the market; massive consolidation and shakeout; low wages and astronomically high turnover; and reliance on selling skills that are fairly general and therefore quite portable across employers. Men’s Wearhouse hires relatively disadvantaged entry level sales personnel, with no prior sales experience (and many with no prior work experience at all). Indeed, a case study of the company (O’Reilly and Pfeffer, 2000) quotes Charles Bresler, Ph.D., the firm’s Executive Vice President of Marketing and Human Resources, as saying:

The retail worker in the United States is somebody who often came from a dysfunctional home... somebody who didn’t do well in school, who basically told the teachers in one way or another to go to hell.

Whereas most competitors rely extensively on temporary and part-time employees and close surveillance, Men’s Wearhouse invests massively in training and inculcating its employees, providing various benefits and a strong team culture that are virtually unheard of in the industry (O’Reilly and Pfeffer, 2000). The result is an uncharacteristically high level of commitment and loyalty to the company, and to its revered (and, until recently, oft-televisioned) founder, George Zimmer.

Why would Zimmer have made such a sizable investment in developing human capital among workers who are untested and risky, in an industry with extraordinarily high turnover (and employee theft), and involving sales and managerial capabilities that are fairly transportable across companies? My contention in this paper is that the psychological bang for the buck here is huge precisely because what is being done here is so atypical, involves heretofore unproven (and sometimes down-trodden) workers, and seems so at odds with a rational employer calculus. Leaders who are fighting “the war for talent” come across as hard-nosed self-interested capitalists: leaders who altruistically embrace overlooked and underprivileged segments of the labor force come across as caring and conscientious citizens.

The goal of this paper is merely to explore the plausibility of this alternative view—whether/when motivational effects produced by showing generosity to non-stars might be of sufficient magnitude to offset pronounced ability differences between stars and non-stars. To do so, I draw on various relevant social science research and reanalyze data from several prior economic studies of gift exchange in labor markets.

2. Arguments for stars

At least three different theoretical rationales exist for when and why employers might focus their recruitment and reward efforts on a relatively small cadre of employees at the top of the distribution of ability and/or effort. First, as Jacobs (1981) has argued, within some roles an extraordinarily high level of performance by a very small group can have a disproportionately large positive effect on overall organizational performance (see also Baron & Kreps, 1999, Chapter 2). This impact is reflected in human resource practices that filter the wheat from the chaff and allocate extremely high compensation (e.g., pay, promotions) to proven stars.9

A second argument emphasizes the value of star talent to organizations as an external signal of product or service quality (Podolny, 2005). Top tier consulting and investment banking firms insist on hiring from elite institutions in part because they believe that clients and prospective employees judge the firm’s attractiveness largely by the pedigree of its talent. As with signaling equilibria generally, this can become a self-fulfilling and self-sustaining belief system: it is not possible for stakeholders to disconfirm their belief in the association between pedigree and performance unless they can observe and relate variance in one to variance in the other.

A third—and related—argument for why organizations may gravitate toward stars is suggested by Frank (1984, 1985) in noting the tendency for reward distributions in organizations to be more compressed than the distribution of contributions or outputs. Frank argues that this reflects the utility that high performers receive in the form of higher local status and the compensating differential that must be provided to the low performers to offset the disutility of their inferior status position:

[W]hen wage schedules are less steep than the standard textbook wage schedule, there results a clear, positive relationship between a worker’s status in the income hierarchy of his firm and the extent to which his wage understates his marginal product. ...One’s status within one’s earnings hierarchy emerges as a good like any other that is traded in the marketplace (Frank, 1984, p. 551).

Frank suggests that the price high status workers are willing to pay for their status will depend, among other things, on the extent of their interaction and interdependence with the rest of their coworkers, through which they are enabled to consume their superior status. Low status workers will be compensated at a price that reflects the utility they could achieve by moving to another firm where they would have higher standing and thus could consume more status. Frank’s argument seems to imply that if an employer can attract and retain stars who care enough about their status vis-à-vis co-workers (or can be made to care), the firm may be able to purchase stars’ services for a considerable discount relative to their marginal product.

9 However, Groysberg (2012) analyzed performance and career mobility among Wall Street equity research analysts, documenting that star performance is not portable across organizational boundaries. This suggests that such extraordinary performances are largely specific to the firm—star match, dependent on systems, culture, and worker interdependencies within the organization, rather than a reflection of fixed individual endowments or abilities.
This would provide another rationale for employment regimes favoring stars.6

Each of these arguments no doubt has relevance to employment relationships in particular contexts. What each overlooks, however, is the possibility that employees’ marginal products may be affected significantly not only by the amount of value they derive from their compensation (pay, status, etc.), but also by the value they derive from the manner in which those rewards are allocated. In other words, employee behavior and attachment will be affected by the extent to which the perquisites being offered are construed as compensation earned versus as discretionary gifts that convey something about how the employer views the employee and the employment relationship.7

The efficiency wage hypothesis focuses on prospects for increasing employee output, loyalty, self-management, and the like by making it costly for the worker to contemplate losing his or her current employment. What I have called the empathy wage hypothesis emphasizes prospects for achieving those same outcomes by making it costly for the employee to contemplate losing his or her current employment relationship, where that relationship is viewed as having value apart from the tangible rewards and benefits it conveys. I argue below that firms have much more limited prospects for benefiting from paying such empathy wages to employees who are viewed (by themselves and/or others) as stars versus employees lower down in the distribution. This potentially can offset or undermine some of the putative advantages of hiring stars mentioned above.

6 It is worth noting that other behavioral logics, no less plausible than those posited by Frank, might imply quite different labor market dynamics. Status reflects not simply one’s relative position in a relationship or social entity, but also the status of that relationship or entity in relation to others. Consider two young aspiring actors, working as personal assistants to two entertainers: Tom Hanks and Pee Wee Herman. Frank’s argument seems to imply higher compensation for Tom’s assistant than Pee Wee’s, to compensate for the larger gap in social standing between the star and the assistant. Yet Tom’s assistant no doubt consumes status and enjoys opportunities by virtue of his/her association with a more revered personality, suggesting that many more people would wish to work for Tom than for Pee Wee, driving up the wages for Pee Wee’s assistant relative to Tom’s.

Indeed, lower productivity workers may derive psychic or material benefit (reflected status, access to informal coaching, etc.) by associating with higher productivity workers in the same organization. And instead of creating feelings of inferiority that must be offset by a steeper earnings profile, low performers’ association with highly productive workers might motivate the former to work harder, because: (a) they wish to reduce the conspicuous gap in productivity and social standing; and/or (b) their association with high producers causes the low performers to raise their self-assessment.

7 Baron and Kreps (1999, p. 110) offer an example of how the same perquisite can come to be viewed very differently, depending on employer actions, involving the provision of unlimited free beverages to employees of Cisco Systems during the company’s start-up phase. After a few years, newly hired purchasing professionals attempted to achieve cost savings by modestly reducing the vast array of beverage choices being provided. Employees revolted and crashed the company’s servers, stalling all business operations. Ultimately, the firm relented and reinstated the status quo ante. However, now it was clear to the employees that someone was calculatively managing the costs of this. Employees no longer construed the beverages as a gift exchange; rather, concern and largesse: rather, it was construed as a form of supplementary compensation that the company had now been browbeaten into providing.

3. Gift exchange and prospects for gratitude

In his classic statement on the norm of reciprocity, sociologist Alvin Gouldner (1960) argued that:

... obligations of repayment are contingent upon the imputed value of the benefit received. The value of the benefit and hence the debt is in proportion to and varies with—among other things—the intensity of the recipient’s need at the time the benefit was bestowed (“a friend in need...”), the resources of the donor (“he gave although he could ill afford it”), the motives imputed to the donor (“without thought of gain”), and the nature of the constraints which are perceived to exist or to be absent (“he gave of his own free will...”). Thus the obligations imposed by the norm of reciprocity may vary with the status of the participants within a society.

Gouldner’s formulation calls attention to at least four dimensions of gift exchange that are likely to affect reciprocity: (a) the perceived value or meaning that the recipient attaches to the donor’s largesse; (b) the recipient’s attributions about the motivations of the giver; (c) the relative status of the giver and recipient; and (d) the nature and history of their relationship. Gouldner proposes that gifts are most likely to elicit reciprocation: among those who are neediest and/or of relatively lower status; when the gift is viewed as having been given with benevolent intent; when the donor’s resources are limited (or perceived as such); when the donor is viewed as not having been coerced or pressured into offering the gift; and when the relationship between giver and recipient has heretofore been one oriented toward affiliation and mutual support rather than tangible instrumental aims of each party. A comprehensive review of helping behavior by Flynn (2006) suggests considerable empirical support for those predictions about how the degree of indebtedness or reciprocation elicited by a favor or gift varies across settings.

These considerations suggest that discretionary gifts may have greater motivational impact when offered to relatively less advantaged or privileged employees and when offered by relatively less prosperous employers and/or those who downplay the role of extrinsic incentives.8

Along these lines, it should be clear that personalization of gifts is likely to increase gratitude and feelings of obligation. Personalization can take the form either of individualizing the content of the gift or the manner of its delivery, so that the recipient perceives it as having been given to him or her as a person, not by virtue of occupying a particular role or organizational category.

Anthropologists and legal scholars have noted that the distinction between gifts and exchanges is a murky one (Mauss, 1954; Rose, 1992). Gifts often have a self-interested purpose, reflected in anthropologists’ references to ‘gift exchange’; conversely, exchange is often undergirded by a gift-like component, such as a willingness voluntarily to forego gain today in order to promote collaboration with a partner over the long term. Quite obviously, we expect employees’ willingness to reciprocate treatment from the employer to be greatest when they highly value what they have been given and view that gift as reflecting benevolent motives and genuine caring. It is less clear how those two dimensions interact, that is, which has the stronger motivational effect: a payout of $X which is clearly an employee compensation as a gift exchange versus a purely self-interested “payment” by the employer of for exchange; versus being “gifted” the proverbial stale fruit cake or hideous cardigan sweater (costing $X) from a donor whose motivations are perceived as entirely benevolent.

8 A comprehensive review of helping behavior by Flynn (2006) suggests considerable empirical support for those predictions about how the degree of indebtedness or reciprocation elicited by a favor or gift varies across settings. These considerations suggest that discretionary gifts may have greater motivational impact when offered to relatively less advantaged or privileged employees and when offered by relatively less prosperous employers and/or those who downplay the role of extrinsic incentives.
When employees have low expectations, inferior social standing, or a prior history of limited attainment, the employer has greater prospects for influencing both the perceived value of and perceived motive for the compensation it provides, in ways that can redound to the benefit of the firm. At the upper end of the talent distribution, in contrast, premium compensation is much more likely to be construed as payment for services rendered than a gift. There is less opportunity for potential employers to frame their offer as anything other than a competitive response to the war for talent, and stars are inclined (indeed, encouraged, by those who fawn over them) to believe that they are worth whatever they can command. Note also that the prospective employers of star talent are very often the dominant firms in their respective sectors; indeed, in industries such as professional services, the elite pedigree of employees is itself one of the strongest bases for an organization’s legitimacy and ability to charge premium rates for its services. As a result, the prospective employers of star talent may be perceived as having more abundant resources, facing stronger market and institutional forces that foster intense competition for top talent, and utilizing a more calculative logic for determining what to offer specific prospects. In some settings (e.g., professional sports, senior executive positions, entertainers), star talent may have formal representation, which certainly strengthens the perception of a market-based relationship between self-interested actors, rather than a heartfelt bond between parties who care about one another. According to Gouldner, being perceived in this way by actual or potential employees is likely to weaken the tendency to reciprocate for superior treatment through increased effort, loyalty, and the like.

In sum, behavioral perspectives on reciprocal altruism and gift exchange suggest that the returns from offering above-average treatment to workers will depend on: how employees value what the employer is offering; employees’ attributions about the employer’s motivations; and the relative cost of higher-versus lower-ability employees. When higher levels of effort elicited through gift exchange can compensate for lower levels of ability, employers may be able to profit by focusing on undervalued human assets, not simply because there is less intense competition for those assets, but also because the employer’s attention causes the assets to think more highly of themselves and of the employer. A variation on this theme was popularized by Michael Lewis’ best-seller *Moneyball* (2003), which described how professional baseball’s obsession with star talent seemingly results in the mispricing of human assets, creating arbitrage opportunities for teams like the Oakland Athletics seeking underpriced non-star talent.10 Interestingly, Billy Beane, the Athletics’ manager and architect of their labor market strategy, describes his approach in precisely these terms:

“ Arbitrage. We don’t use that word too much in baseball, but that’s what it is. In a market where people are competing for scarce assets—for us, it’s players or, really, the things that players can do—there’s always going to be some inefficiency. We’re always going to have to find that dark corner, the stone that hasn’t been turned over” (Hammond, 2003).

Prospects for such arbitrage depend on at least three basic factors: the relative productivity of the two groups; their relative compensation; and, importantly, group differences in responsiveness to incremental units of premium compensation or treatment. The first of these represents what might be thought of as the ability or talent gap; the second captures the compensation gap. I will refer to the third factor as the “gratitude gap.”

Table 1 summarizes the relationship among these three gaps. It shows, for a given level of relative productivity and pay between stars and non-stars, the amount by which gratitude or marginal responsiveness among non-stars must exceed that of stars to leave the employer indifferent between the two groups, all else being equal. For example, if stars are typically three times as productive as the rest of the workforce but only 75% more costly, then marginal responsiveness to incremental rewards or favorable treatment by the employer must be 71.4% higher among non-stars than stars to leave the employer indifferent between the two groups of workers.12

Employees’ responsiveness to premium compensation, as well as an employer’s potential to benefit from efficiency (or empathy) wages, certainly may depend on factors other than those identified in Table 1, which is intended to apply ceteris paribus Efficiency wages for high performers may be justified simply because stars are more costly to replace or less costly to supervise. Moreover, in jobs having what Baron and Kreps (1999, Chapter 2) term a guardian profile, extremely poor performance disproportionately harms organizational performance; conversely, in star roles, extremely high performance generates disproportionate benefit. Efficiency wage regimes are

---

10 Recent research suggests that such mispricing of human assets may be widespread in professional sports. For instance, Massey and Thaler (2013) provide a large body of compelling evidence suggesting that NFL teams substantially overvalue high draft picks (i.e., pay too high a price for early picks in trading draft choices).

11 As noted above, the empathy wage formulation requires not just that non-stars are more grateful or motivated than stars by premium treatment, but also that non-stars can compensate for at least some of the ability gap by exerting greater effort.

12 In this simple example, an average non-star produces 10 widgets per hour for $10, whereas stars average 30 widgets per hour at a wage of $17.50. Suppose that a 10% pay increase to stars boosts their average output by 20%. The average star now produces 6 additional widgets per hour for $1.75 more in wages, or 3.429 additional widgets per dollar of incremental compensation. If the employer raised non-star wages by the same proportion (10%), hourly wages would rise by $1 to $11. To generate the same return per dollar of incremental labor cost, the average non-star must produce an additional 3.429 widgets per hour, which represents an increase of 34.28% over the prior hourly average output of 10 widgets. Hence, when gratitude (i.e., marginal response to incremental compensation) is (34.29/20) = 1.7145 times greater among non-stars than stars, the employer is indifferent between the two groups, all else being equal.
better suited to the latter than the former. On the other hand, several factors work in the opposite direction, potentially limiting the benefits of efficiency wage regimes when applied to stars. When there are human or technological ceilings on performance, for instance, there may simply be less room for incremental output by stars versus non-stars. Given diminishing marginal utility of rewards and increasing marginal disutility of effort, it may be prohibitively expensive to elicit further improvements from an employee who is already displaying superlative performance. The scope for positive signaling or expectations effects by offering premium compensation is also more limited when dealing with star performers, compared to employees lower in the ability distribution. For workers whose ability has not yet been demonstrated or recognized, the firm’s willingness to take the risk of providing premium compensation and benevolent treatment signals a confidence in the employees that they may not previously have had in themselves, raising self-esteem and thereby creating a basis for gratitude on the employee’s part and feelings of obligation to the employer that transcends material concerns.

4. Relative performance and pay among star employees

To gauge the plausibility of the empathy wage hypothesis, it would be useful to have some real-world estimates of differentials in productivity, pay, and gratitude between star and non-star employees, which the following sections of this paper attempt to provide.

One key issue is the relationship between pay and productivity for stars versus non-stars. If stars are actually cheap relative to what they contribute, then generating all the gratitude in the world may be insufficient to justify an arbitrage strategy oriented toward non-stars. What does extant evidence suggest regarding the distributions of performance and pay among stars and non-stars? This section of the paper summarizes some pertinent data from a variety of employment settings. Then, we re-examine data from several prior studies of gift exchange processes, in order to try and gauge the plausibility and potential magnitude of a gratitude gap.

4.1. Performance differentials

4.1.1. Professional baseball

Professional sports is one of the settings where star power is thought to matter most, particularly contexts such as baseball offense, which involves relatively little interdependence among players’ performances. Table 2 presents statistics on offensive performance during the 2009 regular season of Major League Baseball, comparing the top decile of batters on each performance dimension to the remaining 90% of hitters. Home run hitting seems to be the most skewed dimension of offense: during the regular 2009 season, the top 10% of home run hitters on average produced 5.5 times as many homers as the rest of the league, generating 40.1% of all home runs that season. Runs batted in (RBIs) also demonstrates a fairly skewed distribution: upper decile players generated on average nearly four times as many RBIs as the typical player in the bottom 90% of the distribution, and the upper decile was responsible for 29.8% of all RBIs during the 2009 season.

Table 1
Responsiveness to additional pay by non-stars (as percent of stars’ responsiveness) required for break-even return.

<table>
<thead>
<tr>
<th>Productivity multiple (stars:nonstars)</th>
<th>1.05</th>
<th>1.1</th>
<th>1.25</th>
<th>1.5</th>
<th>1.75</th>
<th>2</th>
<th>2.5</th>
<th>3</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative pay multiple (stars:non stars)</td>
<td>1.00</td>
<td>0.79</td>
<td>0.70</td>
<td>0.61</td>
<td>0.53</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
<td>0.25</td>
<td>0.20</td>
</tr>
<tr>
<td>1.25</td>
<td>1.00</td>
<td>0.88</td>
<td>0.79</td>
<td>0.69</td>
<td>0.61</td>
<td>0.53</td>
<td>0.47</td>
<td>0.42</td>
<td>0.37</td>
<td>0.25</td>
</tr>
<tr>
<td>1.5</td>
<td>1.00</td>
<td>0.88</td>
<td>0.82</td>
<td>0.74</td>
<td>0.65</td>
<td>0.57</td>
<td>0.51</td>
<td>0.46</td>
<td>0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>1.75</td>
<td>1.00</td>
<td>0.88</td>
<td>0.82</td>
<td>0.77</td>
<td>0.68</td>
<td>0.60</td>
<td>0.54</td>
<td>0.49</td>
<td>0.44</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.77</td>
<td>0.71</td>
<td>0.64</td>
<td>0.58</td>
<td>0.53</td>
<td>0.48</td>
<td>0.46</td>
</tr>
<tr>
<td>2.25</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.78</td>
<td>0.72</td>
<td>0.65</td>
<td>0.59</td>
<td>0.54</td>
<td>0.49</td>
<td>0.47</td>
</tr>
<tr>
<td>2.5</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.80</td>
<td>0.74</td>
<td>0.67</td>
<td>0.61</td>
<td>0.56</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td>2.75</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.81</td>
<td>0.75</td>
<td>0.68</td>
<td>0.62</td>
<td>0.57</td>
<td>0.53</td>
<td>0.51</td>
</tr>
<tr>
<td>3</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.81</td>
<td>0.76</td>
<td>0.69</td>
<td>0.63</td>
<td>0.58</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>3.5</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.81</td>
<td>0.78</td>
<td>0.72</td>
<td>0.66</td>
<td>0.61</td>
<td>0.57</td>
<td>0.55</td>
</tr>
<tr>
<td>4</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.82</td>
<td>0.78</td>
<td>0.72</td>
<td>0.66</td>
<td>0.62</td>
<td>0.58</td>
<td>0.56</td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.83</td>
<td>0.80</td>
<td>0.74</td>
<td>0.68</td>
<td>0.64</td>
<td>0.60</td>
<td>0.58</td>
</tr>
<tr>
<td>7.5</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.84</td>
<td>0.81</td>
<td>0.75</td>
<td>0.69</td>
<td>0.65</td>
<td>0.61</td>
<td>0.59</td>
</tr>
<tr>
<td>10</td>
<td>1.00</td>
<td>0.88</td>
<td>0.83</td>
<td>0.85</td>
<td>0.82</td>
<td>0.77</td>
<td>0.71</td>
<td>0.67</td>
<td>0.63</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Shaded area in Table 1 represents region in which non-star gratitude must exceed that of stars for break-even return.

To be sure, eliciting what Williamson (1975, p. 69) calls “consummate” cooperation from employees, rather than merely “perfunctory” cooperation, may be valuable even in guardian-type roles. The point is simply that the organizational returns from a marginal increment in effort or vigilance or enthusiasm are likely to be more limited in settings such as guarding a nuclear power plant or flying an airplane, compared to star settings such as big ticket sales or research and development. To avoid such differences and complexities, I focus here on contrasting employee groups doing fundamentally the same work but differing in their average ability or output.

13 To be sure, eliciting what Williamson (1975, p. 69) calls “consummate” cooperation from employees, rather than merely “perfunctory” cooperation, may be valuable even in guardian-type roles. The point is simply that the organizational returns from a marginal increment in effort or vigilance or enthusiasm are likely to be more limited in settings such as guarding a nuclear power plant or flying an airplane, compared to star settings such as big ticket sales or research and development. To avoid such differences and complexities, I focus here on contrasting employee groups doing fundamentally the same work but differing in their average ability or output.

14 Admittedly, some aspects of offense—such as sacrifices, hit and run plays, or the number of runs batted in—depend on the offensive performance of other players on the same team. Whether a batter strikes out, hits a home run, or is hit by a pitch during a given at bat, however, depends very little on the performance of the batter’s own teammates (and much more on how the opposing team’s pitcher and fielders perform).

15 The baseball tabulations are based on the “Sean Lahman Baseball Database” (version 5.7, released December 28, 2009), downloadable at http://baseball1.com/content/view/57/82/. Similar analyses were also conducted for the 2004 and 2005 regular seasons, producing nearly identical results.
The star advantage of top hitters apparently is diminished considerably when assessed by performance measures that characterize average results over considerable time intervals (e.g., batting average), rather than measures of cumulative total counts, such as home runs.\textsuperscript{10} For statistics that measure average overall offensive contribution, such as runs created per at bat and the sum of on-base and slugging percentages, Table 2 indicates that average offensive contribution among the top decile of major league hitters was roughly 140–190% of the average among all other players. The larger differentials reflected in count measures of cumulative results presumably reflect at least in part the fact that such measures do not take into account differences in opportunities, with lower performers in baseball being more likely to be traded, injured, benched, or placed lower in the batting order alongside other mediocre hitters. This is borne out in Table 3: differences between the top decile and all other batters in the number of plate appearances (total offensive opportunities) are considerably larger for count measures (e.g., home runs, RBIs) than for measures that represent performance averages.

4.1.2. Commission sales: major office equipment

Commission-based selling is another setting in which superstar performers are typically believed to generate a very large fraction of sales. Chung, Steenburgh, and Sudhir (2013) recently analyzed the effects of bonuses on sales productivity for account managers of a Fortune 500 corporation that sells durable office products. At my request, they tabulated mean and median monthly sales for the top 10% and bottom 90% of account managers in terms of productivity, which are reported in Table 4.\textsuperscript{17} The pattern is strikingly similar to the baseball results. For total annual revenues, the average salesperson in the top decile generated 5.69 times the revenue of all other

---

\textsuperscript{10} Batting average equals the number of hits divided by “official at-bats” (plate appearances minus [walks + intentional walks + hit by pitch + sacrifices]). On-base percentage equals (hits + walks + hit by pitch) /[official at-bats + walks + hit by pitch + sacrifice flies]; it is regarded as an indicator of a player’s propensity to reach base for reasons other than actions or mistakes of fielders on the opposing team. Slugging percentage is a modified batting average that measures offensive power; it is based on a weighted total of the number of hits, with weights of 1, 2, 3, and 4 for singles, doubles, triples, and home runs, respectively. In recent years, the sum of the on-base and slugging percentages has become quite popular among baseball statisticians (and team managers) as a composite indicator of a player’s offensive contribution. Another frequently used measure of a player’s overall offensive value is runs created (RC), which equals: \((H + BB + CS \times HBP – GDP) \times \left( \frac{TB}{BB + HBP + SF + SH} \right)\), where \(H\) is hits, \(BB\) is walks (base on balls), \(CS\) is caught stealing, \(HBP\) is hit by pitch, \(GDP\) is grounded into double play, \(TB\) is total bases, \(HBP\) is intentional base on balls, \(SH\) is sacrifice hit, \(SF\) is sacrifice fly, and \(AB\) is at bats. Among the hitters we analyze in 2009, RC correlates 0.95 with runs batted in.

\textsuperscript{17} I thank Doug Chung of Harvard Business School and K. Sudhir of Yale School of Management for tabulating these statistics. Total revenues aggregates all sales in 1999–2001 for each account manager and then distinguishes managers who were in the top 10% of that distribution. Average monthly revenues computes each manager’s average monthly sales over that same period, distinguishing the top 10% of producers by that criterion.

---

<table>
<thead>
<tr>
<th>Performance statistic</th>
<th>Top decile</th>
<th>Bottom 90%</th>
<th>Ratio: Top to Bottom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home runs</td>
<td>Mean 29.69</td>
<td>5.39</td>
<td>5.50</td>
</tr>
<tr>
<td></td>
<td>Median 28</td>
<td>3</td>
<td>9.33</td>
</tr>
<tr>
<td>Runs batted in</td>
<td>Mean 99.13</td>
<td>26.56</td>
<td>3.73</td>
</tr>
<tr>
<td></td>
<td>Median 97.50</td>
<td>19</td>
<td>5.13</td>
</tr>
<tr>
<td>Batting average</td>
<td>Mean 0.316</td>
<td>0.225</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>Median 0.310</td>
<td>0.241</td>
<td>1.29</td>
</tr>
<tr>
<td>Runs created per at bat</td>
<td>Mean 0.181</td>
<td>0.095</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td>Median 0.176</td>
<td>0.105</td>
<td>1.68</td>
</tr>
<tr>
<td>Slugging %</td>
<td>Mean 0.543</td>
<td>0.341</td>
<td>1.59</td>
</tr>
<tr>
<td></td>
<td>Median 0.534</td>
<td>0.375</td>
<td>1.42</td>
</tr>
<tr>
<td>On base % + slugging %</td>
<td>Mean 1.102</td>
<td>0.649</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>Median 1.085</td>
<td>0.699</td>
<td>1.55</td>
</tr>
</tbody>
</table>

* Batter's with 20 or more completed at-bats during 2009 Major League Baseball Season.

<table>
<thead>
<tr>
<th>Batting statistic</th>
<th>Average plate appearances</th>
<th>Ratio: average for Top 10% vs. Bottom 90%</th>
<th>Plate appearances: players in Top decile as % of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Home runs</td>
<td>622.79</td>
<td>257.77</td>
<td>2.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.7%</td>
</tr>
<tr>
<td>Runs batted in</td>
<td>642.56</td>
<td>258.12</td>
<td>2.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>22.1%</td>
</tr>
<tr>
<td>Runs created per at-bat</td>
<td>519.45</td>
<td>273.02</td>
<td>1.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>17.3%</td>
</tr>
<tr>
<td>Batting average</td>
<td>462.52</td>
<td>278.62</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13.4%</td>
</tr>
<tr>
<td>On base % + Slugging %</td>
<td>446.68</td>
<td>280.72</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15.1%</td>
</tr>
</tbody>
</table>

* Batter’s with 20 or more completed at-bats during 2009 Major League Baseball Season.
sales people in the company. (Recall that for home runs during the 2009 MLB season, the comparable statistic was 5.50.) As was also true of the baseball statistics, the ratio of star to non-star performance is considerably less skewed when assessed by statistics that characterize average results over considerable time intervals (e.g., average monthly revenues), rather than measures of cumulative total results (total sales revenues).

### 4.1.3. Commission sales: real estate

Real estate is another domain of commission-based sales activity in which elite stars are thought to generate the lion’s share of revenues. Table 5 reports estimates of projected gross commission income in 2009 for residential and commercial real estate: (a) for the industry as a whole in the United States and (b) for “RealDeal” (a pseudonym), one of the industry’s leading firms, which is recognized as having a unique incentive structure to reward top performers. The results for the industry as a whole are rather similar to those reported above for the office products company: on average, the top 10% of real estate salespeople generated about four and a half times more commission income than the rest of the sales force, accounting overall for about one-third of all commission income (compared to 39% in the office products firm). Reflecting its distinctive culture and incentive system, “RealDeal” displays an even more lopsided distribution of results, with a sales force whose output resembles the highly skewed distribution of home runs by major league ballplayers.

#### 4.1.4. Windshield installers at safelite glass

Lazar’s (2000) analysis of Safelite automobile glass installers provides another data point on typical productivity differences between stars and non-stars in settings where worker interdependencies are minimal. Before Safelite experimented with the pay for performance initiatives described and analyzed by Lazar, the bottom 90% of installers averaged 2,423 completed jobs (windows installed) per day, compared to the 5,253 for top decile, implying the latter were on average 116.8% more productive in this setting prior to the adoption of piece rates. After implementation of the performance pay plan, daily output increased to 6,365 for the top decile (+21.2%) and to 2,896 (+19.5%) for the remainder of Safelite’s labor force, implying an average productivity differential of 119.8% between the two groups.

#### 4.1.5. Agriculture: piece rate tree planting in British Columbia

Agrarian production is another context in which performance-based pay is widespread, enabling employers to control unit costs, reward differences in workers’ stamina and physical productivity, reduce the need for direct supervision, and keep employees motivated at tasks

---

Table 4

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Top decile</td>
</tr>
<tr>
<td>Top decile</td>
</tr>
<tr>
<td>Bottom 90%</td>
</tr>
<tr>
<td>Ratio: Top 10% Vs. Bottom 90%</td>
</tr>
</tbody>
</table>

Source: Tabulations provided by Doug Chung and K. Sudhir, based on data analyzed by Chung, Steenburgh, and Sudhir (2013).

Table 5

<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. real estate industry</td>
</tr>
<tr>
<td>Average gross commission income (GCI) per sales associate (2009, projected)</td>
</tr>
<tr>
<td>Top 10%</td>
</tr>
<tr>
<td>Ratio: (Top 10% vs. Bottom 90%)</td>
</tr>
<tr>
<td>Share of all commission income</td>
</tr>
<tr>
<td>Bottom 90%</td>
</tr>
</tbody>
</table>

Source: Statistics tabulated at my request by “Real Deal” firm (which requested anonymity), utilizing internal data as well as proprietary data sources on sales activity within the U.S. real estate industry.
that are repetitive and enervating. Bellemare and Shearer (2009) reported on a field experiment in which a Canadian tree planting firm that utilizes piece rates gave an unexpected, one-time cash gift to its employees. Daily productivity of 18 employees was tracked carefully for approximately 11 weeks before the gift and for one week after, with detailed controls for situational factors (weather, day of week, etc.) that also influence worker output.

Table 6 shows the average daily productivity, prior to the gift, of the top two (11.1%) and top six (33.3%) of the 18 planters studied by Bellemare and Shearer, relative to the rest of the company’s workers.21 The dispersion of performance among tree planters is less pronounced than among the star Safelite installers or groups of salespeople contrasted above: roughly 40–50% above the rest of the labor force.

21 I am very grateful to Charles Bellemare of Laval University for providing the analyses of his proprietary tree planting data that are reported in this paper.

4.1.6. Research Faculty at Cornell University

In documenting the compression of pay relative to the distribution of output, Frank (1984) examined grant-getting by Cornell chemistry faculty (1979–80). He reported that the top 30% generated an average of indirect cost recovery for Cornell of $245,000; the remaining 70% generated an average of $56,814. Thus, the top 30% of grant-getters on average garnered 4.3 times the indirect cost recovery garnered by the bottom 70%.

4.1.7. Summary

Table 7 summarizes the various data sources we have examined concerning the dispersion of performance within various work settings. The magnitude of performance dispersion is fairly similar across disparate contexts: indicators of average performance portray stars as typically outperforming non-stars by 39–117%, with the measures fairly uniformly distributed throughout that range. Measures that simply cumulate results (total home runs, total grant revenues, total sales revenues), and thus do not control for differences in opportunities, produce more highly skewed ratios of star
productivity to non-star productivity, typically between 3.5:1 and 5.5:1.

Bear in mind that most of these data sources involve organizations using piece rates or other types of individual performance-based pay. The fact that such forms of compensation are used suggests that these are settings in which individual differences in ability and effort are believed to be sizable. Accordingly, the data in Table 7 almost certainly represent a (generous) upper bound on stars’ performance advantage: differentials between top performers and the rest of the workforce in these enterprises are likely to be markedly higher than what we would expect to observe in most other kinds of work settings. Moreover, the performance metrics reported in Table 7 are for the most part unidimensional, whereas real-world performance is typically multifaceted, involving some contributions that are more elusive to measure. Where there are trade-offs among performance dimensions—for instance, a Safelite window installer who takes time away from her own piece-rate work to help train a colleague—overall employee contributions are likely to be considerably less dispersed than appears to be the case based solely on the easiest to measure dimension(s) of performance.

4.2. Pay differentials

Frank (1984, Table 6) documented the compression of pay relative to work products in various work contexts. As noted above, among research faculty in Chemistry at Cornell, the top 30% of grant-getters generated 4.3 times more indirect cost recovery for the university than did the remaining 70%. Frank reports that salaries for the former group were only 21.1% higher the following year than the average salary among the remaining chemists, suggesting an extremely compressed pay distribution relative to value presumably contributed to the university through grant-getting. The compression of pay relative to work product is indeed striking in this example. Yet it is difficult to know the extent to which multitasking concerns, inter-temporal considerations, external market forces, and a host of other variables may have shaped the distribution of pay in a particular academic year for career faculty who presumably have other responsibilities to Cornell. Academic institutions are also fairly extreme in terms of pay compression relative to most other kinds of work organizations, a reality captured by the oft-quoted aphorism that politics in academia are so distinctively fierce precisely because the stakes are so abnormally low.

Data from professional baseball, for instance, reveal considerably more pay dispersion than Frank encountered among chemistry researchers, despite performance differentials between stars and non-stars in baseball that look roughly comparable to those Frank documented at Cornell. Using publically available data on compensation for Major League Baseball players, Table 7 compares pay dispersion among star versus non-star batters in 2009 relative to the dispersion of their performance during the same season. Consistent with Frank’s argument, there appears to be somewhat more dispersion in cumulative performance (such as home runs and runs batted in) than in compensation. Although mean performance is roughly 3–4.5 times higher among top decile players, average compensation is about double that of non-stars. But the pattern is by no means uniform; indeed on some measures of performance (e.g., RBIs per at-bat, the sum of on-base and slugging percentage), differences in output between stars and non-stars appear to be smaller than the differences in their respective compensation.

An obvious alternative way of relating pay and performance would be to compare 2009 salaries for stars and non-stars to their relative performance in the preceding season. In such comparisons, we found that performance gaps in 2008 were consistently smaller than the observed dispersion in 2009 compensation (detailed results available on request). For example, among players with available 2009 salary data, the top decile in 2008 produced 34.8% of all home runs, compared to 43.7% in 2009. However, pay dispersion was more pronounced: mean (median) 2009 salary among the top decile of home run hitters in 2008 was 2.66 (4.44) times higher than among the remaining 90%. In sum, for Major League Baseball in 2009, we do not find pay compression as severe as Frank reported, at least based on the relationship between the distribution of salaries and the distribution of players’ offensive performance in the 2008 and 2009 seasons.

What can we conclude from Table 7? Recall that our purpose in pursing these different data sets was to try and calibrate the magnitude of typical differences between stars and non-stars in levels of performance and compensation. That information, in turn, can help us gauge the viability of our empathy wage hypothesis: we can use Table 1 to estimate how much more grateful (i.e., marginally responsive to incremental rewards) non-stars would typically need to be, relative to stars, to offset the relative productivity and cost of the two groups.

In the settings we examined, the top decile of performers outpaced all other employees by roughly 30–120% on measures of average performance; on measures of cumulative or total work product, results for stars typically were about 3–5 times larger than the average among non-stars. Evidence regarding pay dispersion was somewhat less consistent. To be conservative in gauging how big the

---

22 Of course, a small fraction of the very best grant-getters may have captured a disproportionate share of the salary pool, but this seems unlikely in Frank’s sample. For instance, if pay for the top 10% averaged three times that of the bottom 70%, then the remainder of those in the top 30% averaged only 1.1% more than the bottom 70%; if pay for the top 10% averaged twice that of the bottom 70%, the rest of the top 30% must have averaged only 12.2% more than the bottom 70%.

23 For instance, I reanalyzed the 2009 baseball data, dividing batters into the top 30% versus the bottom 70%, to enable direct comparisons with Frank’s results for Cornell research chemists. The top 30% of batters generated 68.4% of all runs batted in and averaged 4.78 times as many RBIs as the bottom 70%, which is very similar to Frank’s findings for the Cornell chemists (see Table 7). The distribution of home runs was a bit more skewed, with the top 30% hitting 77.9% of all home runs in 2009.

24 The data (for 1871–2010), referred to as the “Sean Lahman Baseball Database,” are downloadable from: http://www.baseball1.com.
gratitude gap must be, let us focus on the data points in Table 7 where performance differentials are very pronounced but pay is quite compressed. Looking at professional baseball, for example, the count measures generally indicate performance levels that are 3–5 times higher among stars than non-stars, with pay differentials that are closer to 2:1. According to Table 1, under these conditions of marked performance differentials but fairly modest pay differentials, non-stars must Exhibit 50–150% more gratitude than stars before the employer would be indifferent between the two groups (all else being equal). Obviously, to keep employers indifferent between paying premium wages to non-stars versus stars, less of a gratitude advantage among non-stars would be required if there were smaller differences in productivity or more pay dispersion between the two groups.

5. Is there a gratitude gap? If so, how big might it be?

We know of no data that address the gratitude gap issue unequivocally, so our strategy is to examine several different sources of empirical evidence, from field experiments as well as laboratory studies, which bear on the topic. A recent field experiment among Canadian tree planters provides especially informative data. Bellemare and Shearer (2009, p. 234) summarize their study as follows:

Our study is based on a field experiment, conducted within a tree-planting firm operating in British Columbia, Canada. Workers in this firm are typically paid piece rates and earn approximately $200 per day.

During the experiment workers received a surprise bonus of $80, in addition to their regular piece rate, for one day's planting. The bonus was formulated as a gift from the firm to the workers. Workers were told that extra money was available in the contract due to an exceptional event and that the firm had decided to distribute that money among the workers. We measure worker response in terms of their daily productivity—the number of trees planted.

The experiment was conducted on a large homogeneous block of land permitting the observation of workers, with and without the bonus, under stable planting conditions. Eighteen planters took part in the experiment which took place in the early summer of 2006. The block was planted over a seven-day period and the bonus was paid on the second day of planting on the block. Each worker involved in the experiment is observed planting with and without the bonus. We use our panel data to estimate the effect of the gift on planter productivity, controlling for planter-fixed effects, weather conditions, and other random daily shocks. Our results show that workers responded positively to the gift by increasing their average daily production by 118 trees, approximately 10 percent.

To control for day-of-the-week effects (possibly due to fatigue), we expanded our data set to include information on the daily productivity of the experimental participants over a period of six weeks. This combined data set allowed us to identify the effect of the gift by comparing average productivity on the day of the gift with average productivity both on and off the experimental block, and within and outside of the experimental week. Again, we find that the gift significantly raised average daily planter productivity, by 132 trees, an estimate comparable to that obtained using information on the experimental block alone.

At my request, Bellemare and Shearer generously furnished tabulations from their proprietary data, showing whether/how employees’ responsiveness to the gift depended on their relative productivity prior to the gift experiment. I divided the eighteen planters into three tertiles of six, representing the top, middle, and bottom thirds in average productivity during the six-week observation period prior to the gift. The bottom tertile of employees averaged 648.13 trees per day, compared to 944.70 and 1184.92 for the middle and top thirds, respectively.

Column (1) of Table 8 reproduces the basic results published by Bellemare and Shearer (2009, Table 3, column 2), obtained from regressions predicting daily planter productivity (estimated with robust standard errors). In addition to the variables listed in the table, the analysis controls for: (a) the specific block (plot) being planted, to capture differences in planting conditions that could affect worker productivity; and (b) worker fixed effects to control for stable differences in productivity among the 18 planters. Column (2) adds a control for workers’ average productivity prior to the gift, as well as separate estimates of the effect of receiving the gift among workers who were in the top, middle, and bottom third of the productivity distribution based on their average output during the six weeks prior to the gift. Predictably, worker productivity on the day of the gift was strongly related to average productivity over the preceding six weeks. However, the response to the gift was significantly stronger among workers in the lowest third of the performance distribution than the other two groups. The contrast between the top and middle group of performers is also significant ($b = 45.67; \tau = 2.21, p = .03$).

Fig. 2 portrays these results graphically. The lower portion of each bar in the figure reports the average daily productivity of the top, middle, and bottom tertile of planters prior to receiving the gift. The top set of numbers represents the predicted increment in average daily output for each group on the day of the gift. The model predicts a 35.0% increase in average productivity for the bottom third of workers, a 10.2% increase for the top group, and an increase of 9.3% among the middle third of workers.

---

25 These values are calculated by using the coefficients in column (2) of Table 8 to compute predicted values for workers in the top, middle, and bottom groups. We set productivity to the pre-gift average for each tertile: age, tenure, and the (age × tenure) interaction to their respective sample averages; temperature and rainfall variables to their average values during the post-gift period; and all other variables in the model to zero.
nothing to do with gift exchange and thus deserving consideration:

(1) **Effort ceilings.** Given the intense physical demands of agricultural labor, perhaps responses to the gift were smaller among more productive workers simply because they were already at or near a ceiling on effort or physical work capacity. However, according to Wikipedia, “The average British Columbia planter plants 1600 trees per day, but it is not uncommon for veterans to plant 2000–3000 trees per day while working in the British Columbia interior.” Similarly, a website describing what to expect working as a tree planter in Canada observes that “a single planter does not put hundreds, but thousands of trees into the ground every single day.” Bearing in mind that the top third of workers in this sample averaged only 1185 trees per day, it seems unlikely that high productivity workers were relatively unresponsive to the gift because they were at or near a ceiling on effort or facing physical limitations. Moreover, average response to the gift was at least as strong among the top third of the productivity distribution as among the middle tertile, which is inconsistent with the notion of a ceiling effect.

(2) **Differential relative value of the gift.** Given piece rate compensation, high productivity planters consistently experience higher compensation than their low productivity counterparts. During the six-week period prior to the field experiment, daily piece rate earnings for workers averaged $272.53 for the upper third of workers, $216.50 for the middle third, and $155.97 for the least productive third. The $80 surprise gift thus represented 51.3% of an average day’s pay for the bottom group, compared to 29.4% for the top group. Perhaps low productivity workers responded more because, relative to their income norm, the gift seemed more substantial. However, this explanation does not account for the magnitude of difference in response between the bottom and top groups: as a proportion of daily earnings, the gift was larger by a factor of about 75% for the bottom group (51.3%, vs. 29.4% for the top group); yet the bottom group’s proportionate increase in daily output was 3.5 times that of the top group (35.0% vs. 10.2%, respectively). Nor does this explanation account for why the middle group would have been less responsive than the top group.

(3) **“Shadow of the future”.** Given the seasonality of tree planting work, an alternative explanation for the least productive workers responding disproportionately might simply be a self-interested attempt on their part to increase prospects for returning to work for the same employer during the following planting season. Bellemare and Shearer (2009) examined this possibility for their sample as a whole by interacting the gift effect with whether or not a given worker ended up returning

---

**Table 8**

Regression analyses: response of Canadian tree planters to surprise employer gift.

<table>
<thead>
<tr>
<th>Variable</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day of gift</td>
<td>−28.350</td>
<td>(55.999)</td>
</tr>
<tr>
<td>Day of gift × age</td>
<td>2.639</td>
<td>5.588***</td>
</tr>
<tr>
<td>Day of gift × tenure</td>
<td>94.586</td>
<td>118.877***</td>
</tr>
<tr>
<td>Day of gift × age × tenure</td>
<td>−1.988</td>
<td>−2.547</td>
</tr>
<tr>
<td>Minimum temperature</td>
<td>7.753</td>
<td>7.695</td>
</tr>
<tr>
<td>Maximum temperature</td>
<td>−1.373</td>
<td>−1.391</td>
</tr>
<tr>
<td>Rain</td>
<td>−60.089**</td>
<td>−59.961***</td>
</tr>
<tr>
<td>Minimum temperature × rain</td>
<td>−3.493</td>
<td>−3.481</td>
</tr>
<tr>
<td>Maximum temperature × rain</td>
<td>4.304</td>
<td>4.293</td>
</tr>
<tr>
<td>Tuesday</td>
<td>46.444</td>
<td>46.561</td>
</tr>
<tr>
<td>Wednesday</td>
<td>54.495**</td>
<td>54.444**</td>
</tr>
<tr>
<td>Thursday</td>
<td>54.842**</td>
<td>54.915**</td>
</tr>
<tr>
<td>Friday</td>
<td>−24.962</td>
<td>−25.137</td>
</tr>
<tr>
<td>Average past productivity</td>
<td>0.949</td>
<td>(0.083)</td>
</tr>
<tr>
<td>Day of gift × top third of producers</td>
<td>−90.235***</td>
<td>(22.760)</td>
</tr>
<tr>
<td>Day of gift × middle third of producers</td>
<td>−135.901***</td>
<td>(17.206)</td>
</tr>
<tr>
<td>Day of gift × bottom third of producers</td>
<td>−70.734</td>
<td>(36.401)</td>
</tr>
<tr>
<td>Constant</td>
<td>845.052***</td>
<td>−203.654</td>
</tr>
<tr>
<td>N</td>
<td>623</td>
<td>633</td>
</tr>
</tbody>
</table>

Source: tabulations provided by Charles Bellemare, based on data analyzed by Bellemare and Shearer (2009).

* Significant at 0.10 level (two-tailed).

** Significant at 0.05 level (two-tailed).

*** Significant at 0.01 level (two-tailed).

A similar portrait emerges if worker responses to the gift are estimated based only on the planting data obtained after the gift was given (N = 84). Predicted mean output after the gift for each of the three groups implies productivity increases of 5.8% and 6.5% for the top and middle thirds (respectively), compared to a predicted increase of 25.1% in daily planting among the lowest productivity group.

Employee responses to the gift, in the form of discretionary effort, thus appear to have been markedly stronger among the least productive tree planters relative to other workers. Of course, several alternative explanations for such a finding come immediately to mind, having

---

26 The contrast between bottom and middle groups is significant (b = −87.51; t = −4.08; p < .01, two-tailed); the contrast between the bottom and top groups is marginally significant (b = −60.21; t = −1.99; p < .1, two-tailed). The contrast between the middle and top groups is not significant (b = −27.30; t = 0.71; p > .05, two-tailed).

27 Predicted values are calculated as described in Footnote 25.


the following season; they reasoned that if planters seeking future employment were the ones who responded most to the gift, including that control should eliminate the main effect of the gift.\textsuperscript{30} They reported that planters who returned to the firm the following year did indeed exhibit a stronger response to the gift; however, the effect was nonetheless still significant among the remaining employees as well. We obtained similar results: all else being equal, planters who returned to work for the same firm the following season increased their output after receiving the gift by 155.31 trees more than otherwise comparable employees ($t = 7.86, p < .001$).

In supplementary analyses, we added interactions (to column 2 of Table 8) between productivity tertile and whether or not the employee returned to the same company the following year, thereby allowing the relationship between productivity categories and response to the gift to differ between returning versus non-returning planters. Even among planters who did not return, we observe a (statistically significant) pattern of differential responsiveness as a function of prior productivity: for the least productive third of employees, the model predicts a 14.6% increase in output on the day of the gift compared to their pre-experiment average; in contrast, the middle third is predicted to increase output by 7.6% and output is actually predicted to decrease by 20.9% among the top third (see Table 9). These results suggest that reputational concerns or strategic future-oriented behavior may have affected tree planters at the top and bottom of the performance distribution: in both groups, the workers most likely to exert discretionary effort were the ones who ended up returning to work for the same firm the following year. However, even workers in the bottom productivity group who did not return to the company raised their output markedly in response to the gift, whereas non-returning workers in the top productivity group actually decreased their output substantially.

This evidence on tree planters is certainly consistent with the hypothesis of disproportionate response to gifts by lower performers and less advantaged workers. In that setting, prior to the field experiment, the top third of planters on average were 75% more productive than the bottom third. However, on average, the least productive workers increased their total daily output by 3.4 times as many trees as the top performers (165 vs. 49) and by 10.4 times the average response among the middle third of performers.

6. Psychological bases of reciprocal altruism

Several bodies of evidence and strands of theory from psychology offer insight into how and why reciprocal altruism might be easier to elicit among the non-star, less advantaged segments of the workforce than among the elite stars.

6.1. Attributional processes

Experimental evidence concerning reciprocity and attribution reported by Charness (2004) speaks to how employee responses to employers’ discretionary offers may differ at the top versus the bottom of the wage distribution. Subjects were randomly assigned the role of employer or employee in a laboratory gift exchange game. Charness manipulated not only the wage rate offered to the employee (ranging from 20 to 120), but also whether subjects were told that the wage rate was set: (a) by the employer; (b) randomly; or (c) by a third party (the experimenter). Employees were then asked to record an effort choice (from 0.1 to 1), with the employee’s cost of effort being an increasing function of effort. Payoff functions were common knowledge and subjects divided up their employee and employer shares after completing 10 rounds of the experiment.

Based on data reported by Charness (2004, Tables A1 and B1), Table 10 summarizes how effort varied with wages across the “employer” and “random” conditions of his study, focusing particularly on the extremes of the wage distribution. When employees attributed a low wage offer to employer discretion, their average effort level was reduced by 44% ($0.1000/0.1778 = 0.56$), relative to the effort observed when experimental subjects attributed that same wage offer to a random draw.\textsuperscript{31} The perception

\textsuperscript{30} Note that with person-specific fixed effects included in the model, the main effect of whether or not a worker returned the following year is already controlled for.

\textsuperscript{31} When wages were set by a third party (the experimenter), average effort among the lowest wage subjects (20–29) was 0.13, roughly half way between the effort levels associated with wages set by the employer versus by a random draw.
or attribution that wage decisions reflected an explicit self-interested calculus by the employer seems to have generated a type of “negative gratitude” or resentment on the part of low-wage subjects. Removing that attribution (by assigning wages randomly) boosted effort among low-wage subjects to the point where they became a “bargain”: 32.4% as productive as the high wage subjects, but only 23.3% as expensive.32

6.2. Anchoring and counterfactual reasoning

Psychological research on counterfactual reasoning suggests an additional reason for expecting disproportionate response from less advantaged workers. That body of research has demonstrated how gratitude or satisfaction reflects not simply one’s absolute outcomes, but also how those outcomes compare to what the person might have otherwise plausibly expected. For instance, Medvec, Madey, and Gilovich (1995) famously documented less favorable emotional reactions among silver medalists in the 1992 Summer Olympics and 1994 Empire State Games than among bronze medalists in those same contests, both immediately after the conclusion of the event and while on the medal stand receiving their prizes. They attribute these results to different processes of counterfactual thinking employed by the two groups: silver medalists felt regret relative to the salient counterfactual scenario of having won the gold; bronze medal winners felt gratitude for having made it onto the medal stand, thereby avoiding the salient counterfactual of going home empty-handed. Similarly, Medvec and Savitsky (1997) demonstrated that arbitrary cutpoints powerfully influence gratitude, regret, and satisfaction by influencing the direction of counterfactual thinking: barely surpassing a threshold leaves people feeling more satisfied, whereas falling just short of a threshold leaves people less satisfied. They report that among students receiving a given letter grade (e.g., “B”) in a college course, for instance, those who barely missed receiving the next highest grade are the least satisfied, whereas those who just avoided receiving the next lowest grade are the most satisfied.

The prospect of similar processes affecting gift exchange and reciprocity is suggested by a reanalysis I undertook of data from a prominent experimental labor market study conducted by Fehr, Kirchsteiger, and Riebl (1993).33 Subjects were randomly assigned to the role of employer or worker, interacting anonymously from separate rooms (i.e., trading partners did not know one another’s identities during any trading period) for 12 periods in each session.34

The first stage was a one-sided oral auction with employers as bidders which lasted three minutes. At this stage employers made wage proposals, but they had no opportunity to choose the worker with whom they traded because every worker could accept every offer...If a worker accepted an offered wage p, a binding contract was concluded, and stage 1 was

or attribution that wage decisions reflected an explicit self-interested calculus by the employer seems to have generated a type of “negative gratitude” or resentment on the part of low-wage subjects. Removing that attribution (by assigning wages randomly) boosted effort among low-wage subjects to the point where they became a “bargain”: 32.4% as productive as the high wage subjects, but only 23.3% as expensive.32

6.2. Anchoring and counterfactual reasoning

Psychological research on counterfactual reasoning suggests an additional reason for expecting disproportionate response from less advantaged workers. That body of research has demonstrated how gratitude or satisfaction reflects not simply one’s absolute outcomes, but also how those outcomes compare to what the person might have otherwise plausibly expected. For instance, Medvec, Madey, and Gilovich (1995) famously documented less favorable emotional reactions among silver medalists in the 1992 Summer Olympics and 1994 Empire State Games than among bronze medalists in those same contests, both immediately after the conclusion of the event and while on the medal stand receiving their prizes. They attribute these results to different processes of counterfactual thinking employed by the two groups: silver medalists felt regret relative to the salient counterfactual scenario of having won the gold; bronze medal winners felt gratitude for having made it onto the medal stand, thereby avoiding the salient counterfactual of going home empty-handed. Similarly, Medvec and Savitsky (1997) demonstrated that arbitrary cutpoints powerfully influence gratitude, regret, and satisfaction by influencing the direction of counterfactual thinking: barely surpassing a threshold leaves people feeling more satisfied, whereas falling just short of a threshold leaves people less satisfied. They report that among students receiving a given letter grade (e.g., “B”) in a college course, for instance, those who barely missed receiving the next highest grade are the least satisfied, whereas those who just avoided receiving the next lowest grade are the most satisfied.

The prospect of similar processes affecting gift exchange and reciprocity is suggested by a reanalysis I undertook of data from a prominent experimental labor market study conducted by Fehr, Kirchsteiger, and Riebl (1993).33 Subjects were randomly assigned to the role of employer or worker, interacting anonymously from separate rooms (i.e., trading partners did not know one another’s identities during any trading period) for 12 periods in each session.34

The first stage was a one-sided oral auction with employers as bidders which lasted three minutes. At this stage employers made wage proposals, but they had no opportunity to choose the worker with whom they traded because every worker could accept every offer...If a worker accepted an offered wage p, a binding contract was concluded, and stage 1 was

or attribution that wage decisions reflected an explicit self-interested calculus by the employer seems to have generated a type of “negative gratitude” or resentment on the part of low-wage subjects. Removing that attribution (by assigning wages randomly) boosted effort among low-wage subjects to the point where they became a “bargain”: 32.4% as productive as the high wage subjects, but only 23.3% as expensive.32

6.2. Anchoring and counterfactual reasoning

Psychological research on counterfactual reasoning suggests an additional reason for expecting disproportionate response from less advantaged workers. That body of research has demonstrated how gratitude or satisfaction reflects not simply one’s absolute outcomes, but also how those outcomes compare to what the person might have otherwise plausibly expected. For instance, Medvec, Madey, and Gilovich (1995) famously documented less favorable emotional reactions among silver medalists in the 1992 Summer Olympics and 1994 Empire State Games than among bronze medalists in those same contests, both immediately after the conclusion of the event and while on the medal stand receiving their prizes. They attribute these results to different processes of counterfactual thinking employed by the two groups: silver medalists felt regret relative to the salient counterfactual scenario of having won the gold; bronze medal winners felt gratitude for having made it onto the medal stand, thereby avoiding the salient counterfactual of going home empty-handed. Similarly, Medvec and Savitsky (1997) demonstrated that arbitrary cutpoints powerfully influence gratitude, regret, and satisfaction by influencing the direction of counterfactual thinking: barely surpassing a threshold leaves people feeling more satisfied, whereas falling just short of a threshold leaves people less satisfied. They report that among students receiving a given letter grade (e.g., “B”) in a college course, for instance, those who barely missed receiving the next highest grade are the least satisfied, whereas those who just avoided receiving the next lowest grade are the most satisfied.

The prospect of similar processes affecting gift exchange and reciprocity is suggested by a reanalysis I undertook of data from a prominent experimental labor market study conducted by Fehr, Kirchsteiger, and Riebl (1993).33 Subjects were randomly assigned to the role of employer or worker, interacting anonymously from separate rooms (i.e., trading partners did not know one another’s identities during any trading period) for 12 periods in each session.34

The first stage was a one-sided oral auction with employers as bidders which lasted three minutes. At this stage employers made wage proposals, but they had no opportunity to choose the worker with whom they traded because every worker could accept every offer...If a worker accepted an offered wage p, a binding contract was concluded, and stage 1 was

32 Interestingly, Charness found no evidence of an offsetting effect at the top of the wage distribution: among those receiving the highest pay rate, effort levels were no higher when employees perceived the wage to have been chosen by the employer instead of by a random draw.

33 I am very grateful to Professor Ernst Fehr for furnishing the raw data on which these analyses are based.

34 Sessions 1 and 2 occurred in the morning and afternoon of October 11, 1991; sessions 3 and 4 occurred in the morning and afternoon one week later. Each session lasted 2 h.
Table 11
Determinants of worker effort in an experimental labor market (N = 242).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Range of variable</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>B</td>
<td>t</td>
<td></td>
<td>B</td>
<td>t</td>
<td></td>
<td>B</td>
<td>t</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-.099</td>
<td>-1.449</td>
<td>.149</td>
<td>-.137</td>
<td>-1.162</td>
<td>.246</td>
<td>-.071</td>
<td>-.913</td>
</tr>
<tr>
<td>Wage</td>
<td>30–110</td>
<td>.007</td>
<td>9.900</td>
<td>.000</td>
<td>.007</td>
<td>9.488</td>
<td>.000</td>
<td>.007</td>
<td>10.000</td>
</tr>
<tr>
<td>First wage accepted in current session</td>
<td>10–100</td>
<td>-.003</td>
<td>-4.731</td>
<td>.000</td>
<td>-.003</td>
<td>-2.347</td>
<td>.020</td>
<td>-.002</td>
<td>-3.445</td>
</tr>
<tr>
<td>Effort level in previous period</td>
<td>10–1.00</td>
<td>.460</td>
<td>8.626</td>
<td>.000</td>
<td>.423</td>
<td>7.665</td>
<td>.000</td>
<td>.403</td>
<td>7.368</td>
</tr>
<tr>
<td>Number of dormant periods prior to current period</td>
<td>0–8</td>
<td>.051</td>
<td>2.285</td>
<td>.023</td>
<td>.051</td>
<td>2.324</td>
<td>.021</td>
<td>.046</td>
<td>2.086</td>
</tr>
<tr>
<td>Previous effort × number of dormant periods</td>
<td></td>
<td>-.156</td>
<td>-2.807</td>
<td>.005</td>
<td>-.160</td>
<td>-2.887</td>
<td>.004</td>
<td>-.135</td>
<td>-2.423</td>
</tr>
<tr>
<td>Period within session</td>
<td>2–12</td>
<td>-.002</td>
<td>-.061</td>
<td>.543</td>
<td>-.002</td>
<td>-.588</td>
<td>.557</td>
<td>-.002</td>
<td>-.688</td>
</tr>
<tr>
<td>Adj. $R^2$</td>
<td></td>
<td>0.515</td>
<td>.522</td>
<td>.534</td>
<td>.546</td>
<td>.546</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEE</td>
<td></td>
<td>0.152</td>
<td>0.151</td>
<td>0.149</td>
<td>0.147</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td></td>
<td>43.649</td>
<td>30.195</td>
<td>20.740</td>
<td>18.056</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>df</td>
<td></td>
<td>6</td>
<td>9</td>
<td>14</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects: sessions</td>
<td></td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fixed effects: sellers</td>
<td></td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: GLS estimates from data originally collected and analyzed by Fehr et al. (1993) and generously provided to author by Professor Ernst Fehr.

finished for both the worker and the employer. If an employer’s bid was not accepted, he was free to change his bid, but the new bid had to be higher than the previous highest bid (possibly from other employers) which had not yet been accepted. After three minutes the market was closed, and those parties who did not succeed in trading earned zero profits for this period. At the second stage, workers had to choose their effort anonymously; i.e., their choice was revealed only to “their” employer. Moreover, their choice was completely unconstrained in the sense that there were no sanctions associated with it.

In total, we organized four experimental sessions. In all sessions we deliberately created an excess supply of workers (Fehr et al., 1993, pp. 439–40).

Workers chose an effort level between 0.1 and 1.0, with the cost of effort increasing nonlinearly with effort. The payoff functions to workers and employers were common knowledge to both parties, and subjects were required to demonstrate their ability to compute payoffs for themselves and their trading partners prior to the start of the experimental period. On average subjects earned around 296 Austrian Shillings (~$25 in 1991; roughly $40 in 2010 dollars) per 2-hour session.

Table 11 summarizes linear regression models predicting the level of effort between 0.1 and 1.0 selected by workers in each period. (Results were unchanged by a probit transformation of the effort variable to deal with floor and ceiling effects; details available from author upon request.) Model 1 regresses effort level on the wage offered and the workers’ level of effort during the prior period. To capture any effects of “fatigue” and/or prior “unemployment,” the model controls for the number of dormant periods (i.e., in which the worker was not matched to an employer) immediately preceding the current period within the current experimental session. The model also includes a product term for the interaction of that variable and prior effort level. As an additional control for within-session trends, model 1 includes a linear term representing the sequence of periods (1–12) within an experimental session. Table 11 also reports specifications that supplement model 1 by adding fixed effects for experimental session (model 2), worker (model 3), and both session and worker (model 4).

Psychologists’ theories of counterfactual reasoning imply that a worker’s willingness to exert discretionary effort at a given wage is likely to depend not simply on that current wage, but also on the worker’s wage history, which will influence his or her implicit assessment of how much worse or better things might have turned out. Consequently, to examine potential path dependence in workers’ responses to wage offers, model 1 also controls for the first wage received by the worker during each of the four experimental sessions conducted by Fehr and his colleagues.

Because we are interested in the dynamics of how specific workers responded across trials within experimental sessions, we focus our attention on models 3 and 4 in Table 11, which incorporate worker-specific fixed effects. We replicate the main result reported by Fehr et al. (1993): worker effort increases with the wage. Fehr and his colleagues interpret this as evidence of fairness or gift exchange effects, given that workers could (anonymously and with impunity) have simply selected the minimum effort level of 0.1 after accepting a wage offer, which is what conventional economic theory predicts that self-interested, effort averse workers would do. Accord-
Table 12
Predicted effect of current and starting wage on effort in an experimental labor market.

<table>
<thead>
<tr>
<th>Wage during present period</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>70</th>
<th>90</th>
<th>100</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>First wage received during session</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.202</td>
<td>0.276</td>
<td>0.350</td>
<td>0.497</td>
<td>0.645</td>
<td>0.719</td>
<td>0.793</td>
</tr>
<tr>
<td>70</td>
<td>0.131</td>
<td>0.205</td>
<td>0.279</td>
<td>0.427</td>
<td>0.574</td>
<td>0.648</td>
<td>0.722</td>
</tr>
<tr>
<td>100</td>
<td>0.060</td>
<td>0.134</td>
<td>0.208</td>
<td>0.356</td>
<td>0.503</td>
<td>0.577</td>
<td>0.651</td>
</tr>
</tbody>
</table>

Source: Secondary analyses of data originally collected and analyzed by Fehr et al. (1993) and generously provided to author by Professor Ernst Fehr. Predicted values computed based on model 3 of Table 11 (see Footnote 39).

This pattern of results suggests that subjects’ initial experiences within the experimental labor market are shaping processes of counterfactual thinking, which in turn influence the gratitude or reciprocity elicited by subsequent wage offers. Having experienced a fairly adverse result at the start of a session prompts continued downward counterfactual comparisons throughout the session, apparently eliciting more gratitude at a given level of wages, reflected in more discretionary effort. In contrast, for workers whose initial labor market outcomes were favorable, the salient counterfactual presumably is more likely to be the superior (more lucrative) state of nature they have already experienced, which serves to undermine gratitude and reciprocity.30

The psychological research on counterfactual reasoning thus provides an additional rationale for why gift exchange might plausibly trigger more gratitude among those who start off being relatively less fortunate, talented, or privileged within a distribution. If those employees can readily and credibly envision a dreadful state of affairs that might have existed but for the opportunities provided by the employer, then they may be more grateful, despite having gotten less. Conversely, more advantaged counterparts may have fared much better yet feel less gratitude or satisfaction, because they can easily envision an even better counterfactual outcome.

7. Implications for efficiency wages and gift exchange

Our cursory survey of various sources of data on the distribution of productivity, pay, and “gratitude” was intended as an effort to gauge whether the empathy wage hypothesis seems plausible in light of real-world empirics. Nothing unearthed in our survey suggests the hypothesis should be dismissed ex ante based on extant empirical evidence. Indeed, we found evidence suggesting that:

(1) Responses to unexpected gifts are considerably stronger toward the bottom of the productivity distribution than at the top.
(2) Attributional processes relating to wage setting seem to be more powerful among low wage workers than high wage workers, implying that the former may be

---

38 Obviously, fixed effects for every (subject × session) combination would completely absorb the effect of session starting wage, which only takes on one value per session per subject. Model 4 includes eight fixed effects for subjects and three for sessions, which apparently capture a large fraction of that variation.

39 We fixed the prior effort level at 0.4, the number of dormant periods at 0, and the period at 7, which equal the observed sample medians. All other variables in model 3 of Table 11 were set to 0.

40 Our results are also consistent with psychological research documenting that satisfaction with a dynamic outcome depends not only on the initial value one receives and of (absolute) change in value, but also on the rate of change and change in the rate (Hsee and Abelson, 1991; Hsee, Salovey, and Abelson, 1994).
Table 13
Attributes or dimensions of gift exchange affecting the likelihood of reciprocity.

<table>
<thead>
<tr>
<th>Attribute/dimension of gift</th>
<th>Likely sustainability or persistence of reciprocity</th>
<th>Rationale/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neediness of recipients</td>
<td>Low to moderate</td>
<td>Employees become relatively less needy each time gift is given</td>
</tr>
<tr>
<td>Unexpectedness</td>
<td>Low to moderate</td>
<td>Each gift becomes less surprising than the previous one</td>
</tr>
<tr>
<td>Magnitude of gift, relative cost to donor; degree of difficulty</td>
<td>Moderate</td>
<td>Persistence of reciprocity will depend on how/why employer’s financial standing is perceived to change over successive rounds of gifts</td>
</tr>
<tr>
<td>Benevolence intent (transcending self-interest)</td>
<td>High</td>
<td>Depends on employer avoiding appearance of applying any cost-benefit calculus</td>
</tr>
<tr>
<td>Personalization; gift reflects more than market exchange relation</td>
<td>High</td>
<td>Non-monetary and customized gifts are expected to generate greatest reciprocity</td>
</tr>
<tr>
<td>Volition; lack of coercion, pressure to give</td>
<td>Moderate</td>
<td>The precedent established by giving a gift may tend to make employees regard continuation of the practice as less and less “voluntary,” weakening the propensity to reciprocate</td>
</tr>
<tr>
<td>Gift signals employer's perception of employee potential or ability</td>
<td>High</td>
<td>Employee’s own self-image becomes tied to reciprocation through increased output, loyalty, effort, etc.</td>
</tr>
<tr>
<td>Gift signals employer’s adoption of higher standard of employee treatment</td>
<td>Moderate</td>
<td>Persistence of reciprocity depends on: (a) not losing ground on any other dimension(s) of how gift is construed; and (b) competitors not responding</td>
</tr>
<tr>
<td>Gift encourages downward counterfactual comparisons by recipients</td>
<td>Moderate</td>
<td>Same as above. (Adverse counterfactual comparisons are likely to elicit less gratitude if gift comes to be construed as less benevolent, costly, voluntary, etc.)</td>
</tr>
</tbody>
</table>

more responsive to employer treatment that markedly exceeds (or falls short of) basic worker expectations.

(3) How people respond to dynamic outcomes is powerfully affected by their starting position and by the referent level against which they are assessing outcomes, as well as the trajectory they experience. In particular, individuals whose wage trajectories evidence large improvement after having started at a low initial wage persistently provide more discretionary effort than counterparts who are being paid the same current wage but have not experienced much or any improvement over time.

These findings are especially noteworthy given that none of the empirical studies we drew on was explicitly designed to test for differences in employees’ responsiveness to employer generosity or stinginess.

We have argued that predictions about the returns from “efficiency” or “empathy” wages—and about the likelihood of eliciting reciprocity through gift exchange—rely implicitly or explicitly on assumptions about the psychological or sociological bang for the buck, assumptions that deserve much closer attention in both theoretical and empirical work. We need to understand much better when and why employees feel compelled to reciprocate better-than-expected treatment through discretionary effort or loyalty.

These considerations have been largely absent from both field and laboratory experiments designed to gauge the scope, magnitude, and significance of gift exchange processes in employment relationships (for exceptions, see Charness, 2004; Kirchler, Fehr, & Evans, 1996; Kube, Maréchal, & Puppe, 2012; Stanca, Bruna, & Corazzini, 2009), though they have started to receive more attention from theorists (e.g., Danthine & Kurmann, 2006; Dodlova & Yudkevich, 2009; Dur, 2009; Prendergast & Stole, 2001; Rabin, 1993; Ruffle, 1999). In describing recent laboratory work by economists that explores “conditions that facilitate or weaken the strength of gift exchange,” Gneezy and List (2006, p. 1379) highlight work examining the role of sanctions, whether a comprehensive payoff table is available to all experimental subjects, and the role of “parameterization and implementation considerations” in determining student behavior in experimental labor market games. These extensions are commendable, but our argument and survey of empirical evidence suggests that a much broader and richer set of contextual considerations warrant attention.

Toward that end, Table 13 offers some cursory predictions about dimensions of gift exchange regimes that are likely to elicit reciprocity and how persistent or sustainable that reciprocity is likely to be, based on extant psychological and sociological literatures. An extensive body of work on psychological contracts (e.g., Rousseau, 1995), for example, suggests that employees’ loyalty, felt obligation, flexibility, and willingness to go beyond the minimum effort required is more likely to arise in “relational contracts” (employment relations characterized by longer duration and less precisely specified performance criteria) than in “transactional contracts” that are of shorter duration and/or more specific about performance terms. We suggest that gifts are also likely to elicit sustained reciprocity when they are perceived by recipients as benevolent, personalized (reflecting a relationship that transcends a generic market logic), and where the offer of the gift is construed as a signal of the employer’s confidence in employees’ capabilities. Reciprocity is also likely to increase as a function of: how needy the recipients are; how unexpected the gift is; how voluntary and costly the gift is perceived to be; how clearly it signals the employer’s commitment to a higher standard of employee treatment than is customary; and how easily and vividly employees can engage in downward counterfactual comparisons that highlight worse situations in which they might have found themselves. For reasons specified in the rightmost column of the table, we suspect that these latter dimensions may be less sustainable as sources of ongoing reciprocation.

A major motivation for this paper is to encourage theoretical and empirical work investigating conjectures
about when gift exchange regimes are most likely to evoke sustained reciprocity. We suspect that research along these lines would illuminate a number of issues that currently are unresolved and contested among scholars. For instance, in their field study of how tree planters responded to a one-time employer gift, Bellemare and Shearer (2009, p. 243) documented increased productivity, but found that “the generated response [was] ... short term—no significant effect was found on subsequent planting.” The factors listed in Table 13 might help explain why reciprocity deteriorated rapidly in their study. The $80 gift received by the Canadian tree planters exhibited few of the attributes that the table identifies as likely sources of enduring reciprocity. The gift was small and not very costly to the employer (framed as an unexpected windfall, rather than as a difficult sacrifice by the company’s leadership). It was purely monetary, unclear in intent, and identical for all workers. Nothing in the cover story communicating the gift suggests that employees were likely to have construed the gift as signaling that the employer had raised its assessment of employee capabilities or its standards regarding how employees deserve to be treated.

In the same vein, Gneezy and List (2006) reported that gift exchange effects were extremely transitory in two separate field studies they conducted (also see Burger, Misa, Lisa, Kris, and Christopher, 1997). Their first field experiment was conducted at a large university, inviting undergraduates via posters to assist in a one-time project computerizing the holdings of a small library at the university. Subjects were informed the project would last six hours and that they would receive $12 per hour. Participants receiving the gift treatment were informed, after the task was explained, that they would receive $20 per hour rather than the advertised $12 rate. Participants were seated in front of a computer terminal next to boxes filled with books and asked to enter data regarding the books into a database on the computer. Their second field experiment invited students to participate in a door-to-door fundraising drive to support a charitable endeavor at a large university. Interested parties were told that this was one-time work for which they would be paid $10 per hour. After undergoing training, solicitors in the gift treatment condition were informed that they would receive $20 per hour, instead of the previously advertised $10.

Gneezy and List observed significantly more effort among those receiving the gift treatment in both settings (data entry productivity and donations raised). However, this effect waned quickly; after the first few hours, the gift and no gift groups were statistically indistinguishable. Assessing these two experimental contexts against the dimensions in Table 13 suggests why Gneezy and List might have observed little sustained reciprocity. In both experiments, subjects were college students being recruited to a one-time project. Apparently the compensation initially offered had been sufficient to prompt their participation. It seems unlikely that participating students were highly needy or inclined toward immense gratitude for having received this unexpected opportunity. Given the idiosyncratic and unfamiliar nature of the tasks, students are unlikely to have had a clear sense of how the pay being offered compared to a market wage for such activity and thus how generous the employer’s offer was. Indeed, given the short-term nature of the projects, students may well have assessed the opportunity cost of participating not in terms of income attainable through other kinds of employment, but rather in terms of the value of leisure or study time. Nothing in Gneezy and List’s cover story or description of the experimental procedures suggests that students receiving the higher wage were given any explanation of: why they were receiving it; whether the added expense imposed any kind of burden on the project sponsor; or whether the boost in pay rate was being provided to all participants or only to select individuals. In sum, the experimental settings created by Gneezy and List reflect few of the situational, attributional, or perceptual elements that we expect to promote sustained reciprocity. Accordingly, it seems likely their subjects construed the higher-than-expected wage not as a gift, based on some normative standard, but instead as a windfall, relative to a vague reference level established by the initially advertised pay rate.

8. Implications for research

The conjectures in Table 13 suggest some promising directions in field, survey, and experimental research on employment relationships and wage determination. We have suggested that the likelihood of employees reciprocating above-market treatment is likely to depend on their social standing and prior labor market histories, norms that govern how those employees customarily expect to be treated, the perceived costliness to the employer of the discretionary treatment, and employees’ perceptions or attributions regarding the employer’s motivations. The Men’s Wearhouse example mentioned in the introduction to this paper illustrates a powerful empathy wage regime targeted at relatively less fortunate potential employees, who, given their career histories and the nature of the retail industry, harbor low expectations regarding treatment by and loyalty to their employer. Founder and former CEO George Zimmer conspicuously made numerous costly investments in his employees, despite a high risk that employees could parlay those investments into career opportunities in competing firms. According to accounts by those familiar with the company, an important basis of the bond that employees feel to Men’s Wearhouse is not simply that the company took a risk and showed compassion in giving them a chance at a career, but how this empathic and benevolent treatment is so strongly personified by and instantiated in the firm’s leader, George Zimmer (O’Reilly and Pfeffer, 2000). Indeed, Zimmer appears to have gone to considerable lengths to ensure that his investments in Men’s Wearhouse personnel were not perceived as reflecting a self-interested cost-benefit calculus. A business school case study on the company quotes him as saying:

You know, if you ask me how I measure the results of my training program, I can’t. I have to do it on blind faith and trust in the value of human potential (Pfeffer, 1997, p. 4).
Recent analyses of other kinds of economic exchanges have illustrated similar processes at work in creating arbitrage opportunities. For instance, in a study of microcredit in Mexico, Canales (2013) reports that “particularistic” loan officers, who bend rules and develop close personal relationships with clients, outperform their more universalistic, bureaucratically minded coworkers within the same branch. This is true despite the fact that the former take on a riskier clientele in terms of prospects for loan repayment and business solvency, often providing customized restructuring of bad loans, rather than following the standard prescribed practice of referring such loans to collection. Canales reports that this willingness of particularistic loan officers to take a risk on unspecified borrowers, under girded by a close personal relationship that manifests a genuine concern for the client’s well-being, is what elicits reciprocity from clients, enabling those loan officers to garner superior returns. By bending rules, taking a personal risk, building a close personal relationship with the client, and devising personalized approaches to the predicament of individual clients, particularistic loan officers are able to signal more powerfully: (a) how costly, difficult, or personally risky it is for them to accommodate the struggling client and (b) that their desire to see the client succeed transcends their narrow self-interest in complying with standard operating procedures.

More research along similar lines in the context of employment relationships would be invaluable. Both field and lab studies typically categorize individuals into homogeneous categories (e.g., “worker”) and examine the outcomes individuals attain according to some standard payoff schedule. In reality, however, employers often have considerable leeway in how they can personalize the classification and treatment of individuals, the specific rewards made available, and the criteria for attaining those rewards. Our prediction is that prospects for eliciting reciprocity for employees will be greatest when such personalization of employment relations is costly to the employer, perceived by employees as reflecting benevolent motives, and confers more status or dignity on employees than they expect or are accustomed to. Consistent with these conjectures, Kube et al. (2012) recently demonstrated that the nature of a gift, not its monetary value, is what determines whether it elicits reciprocity. Unexpected gifts-in-kind increased productivity among their experimental subjects markedly and significantly; equivalent cash gifts, on the other hand, were largely ineffective in triggering reciprocal altruism, despite the fact that experiment subjects reported strongly favoring the gift’s cash equivalent over the gift-in-kind.

If employee perceptions of employer motivations are important, the impact of a given organization’s reward regime is likely to depend significantly on what competitors are doing and what the focal organization itself has done in the past. Table 13 suggests that employee propensities to reciprocate favorable treatment depend not only on the perception that the employer’s motives are benevolent, but also that the employer is acting purely voluntarily. The more prevalent a specific offering becomes within an industry, the more that external pressures were perceived to drive the firm’s initial provision of the gift, or the longer-lived that pattern of gifting is within the organization, the more likely employees are to question whether the firm’s continued willingness to provide the gift is truly voluntary, rather than a practice mandated by virtue of industry practice, pressure from external constituencies, or the force of past precedent within the firm. This suggests that the impact of providing above-market wages and benefits may depend on life cycle effects as well as on the initial conditions that prevailed when a specific practice was adopted.

Recent research suggests that employee responses to wages may also depend on cognitive processes triggered by the form of payment. Studies utilizing various methodologies and types of empirical data have recently documented that when time is made salient via the form of payment received (e.g., hourly wage, requiring individuals to log hours worked, etc.), individuals: (a) are more willing to trade time for money (i.e., forego leisure); (b) place greater weight on income in assessing their overall happiness; and (c) are less likely to volunteer, donate to charitable causes, or be willing to do work without pay (DeVo & Pfeffer, 2007a, 2007b, 2009, 2010; Pfeffer & DeVo, 2009). Indeed, simply having a random sample of individuals translate their salary into an implied hourly rate was sufficient to produce comparable effects (DeVo & Pfeffer, 2007a).

This line of research suggests that compensation systems and organizational practices that focus attention on the temporal basis of compensation may promote a calculative orientation toward work, which in turn is likely to undermine intrinsic motivation and feelings of loyalty or obligation to the employer. In other words, activating gift exchange requires not only that employees believe the employer is eschewing a self-interested, cost-benefit calculus, but that employees also perceive themselves as doing likewise. Employee self-perceptions can be influenced profoundly by various employer choices, such as: the form of worker payment; benefits offerings; organizational culture; the extent of reliance on monitoring, close supervision, and formal rules; recruitment and selection processes that de-emphasize extrinsic rewards and require extensive displays of voluntary commitment from prospective hires; and the like.

Groups of employees are also likely to vary in their responses to above-market employer treatment as a function of reference points and processes of counterfactual thinking that shape their evaluations of well-being. Imagine two airline passengers on the same overbooked flight. Had one more seat been available in first class, Jack would have been next on the list for a free upgrade from economy class. In contrast, Jill’s was the very last name on the list of economy-class passengers whose reservations were honored, with everyone below her having been bumped and rebooked on a flight the following day. Psychological research predicts that Jack and Jill will experience the flight (and airline) very differently, given the counterfactual each is likely to invoke in imagining what might have been. In the same vein, employers may be able to cultivate more intense feelings of gratitude and loyalty among employees who believe they narrowly escaped a worse fate than among those who are preoccupied with how much better things could (or should) have been. This suggests that reciprocity
may be more forthcoming among groups of employees who: have historically been relatively disadvantaged; have lower standards or expectations regarding the treatment to which they are entitled; are atypical of the kinds of individuals generally hired in a given setting; or face particularly inhospitable labor market conditions (industry downturns, etc.).

One factor likely to affect the reference points and counter-factuals that employees invoke in gauging well-being is the organization’s dispersion in rewards and treatment. Wide vertical dispersion in pay and other rewards has been shown to reduce loyalty, commitment, and individual and organizational performance, especially in interdependent settings (e.g., Bloom, 1999; Grund & Westergaard-Nielsen, 2008; but cf. Simmons & Berri, 2011). These effects appear to be strongest among those located toward the bottom of the organization’s salary structure (Pfeffer & Langton, 1993). Excessive pay dispersion within organizations might well undercut gratitude elicited via empathy wages, inducing those at the bottom of the pay distribution to shift their counter-factual away from the inferior outcomes they might have endured had they not been employed to the superior outcomes they might have experienced had they been treated as well as those at the top.

In sum, our understanding of employment relationships and labor markets stands to benefit enormously from survey, field-based, and experimental research that plays closer attention to the factors that influence the propensity of employees to reciprocate favorable (or unfavorable) treatment from their employers.

9. Conclusion

The much-ballyhooed “War for Talent” has focused attention on star employees—the very best and brightest, who are viewed as indispensable to success. We undertook a cursory survey of disparate data sources concerning how productivity, pay, and propensities to reciprocate vary among employees within work settings. Differences in output between star (top 10%) and non-star employees typically ranged between 40 and 115%. At the same time, several studies we reviewed suggested that the capacity for employers to elicit feelings of gratitude (or resentment) may be strong enough among non-star and less advantaged employees to more than compensate for even a marked talent disadvantage vis-à-vis stars.41

As reflected in the quotations at the beginning of this paper, stars may have much more to give, but they also expect much more. Despite evidence regarding the contextual underpinnings of stellar performance (Groysberg, 2012), stars are naturally inclined to attribute their star performance to dispositional characteristics, making them believe they are worth whatever they can command. Given intense attention to and competition for stars, their employment relationships are highly prone to have the flavor of pure market exchanges, leaving less room for loyalty, reciprocity, or other bases of psychological leverage from offering premium compensation or treatment. We suggested that processes of counterfactual thinking, studied extensively by psychologists, are likely to focus the attention of stars on how much better things might have been, whereas foot soldiers can more easily imagine how much worse they might have fared.

If correct, our argument suggests the potential for labor market arbitrage strategies that target empathy wages and gift exchange regimes at relatively unsung segments of the labor force. Assessing our argument will require greater attention to the sociological and psychological underpinnings of gift exchange, which we believe will illuminate our understanding of labor markets immensely. Drawing on relevant work from psychology, sociology, and anthropology, we have speculated on some contextual factors that are likely to influence how much gift exchange elicits reciprocity and how sustainable such reciprocity is likely to be over time. If this set of concerns figures prominently in future research, we will feel immense gratitude.

References


---

41 Jennifer Brown (2011) makes a different argument about the perils of managing stars alongside non-stars, suggesting that the presence of the former drives down the average performance of the latter. She demonstrates this effect in statistical analyses of panel data from the men’s professional golf (PGA tour), showing that the participation of Tiger Woods has a sizable negative effect on the performance of others competing in that same tournament, and that this effect is not due to players adopting riskier (higher variance) strategies during tournaments in which they are facing a superstar. Groysberg, Polzer, and Ellenbein (2011) report similar findings for Wall Street sell-side equities research analysts; groups benefited—up to a point—from having high status members, controlling for individual performance, but that effect abated and eventually turned negative.


