



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

When $3 + 1 > 4$: Gift Structure and Reciprocity in the Field

Duncan S. Gilchrist, Michael Luca, Deepak Malhotra

To cite this article:

Duncan S. Gilchrist, Michael Luca, Deepak Malhotra (2016) When $3 + 1 > 4$: Gift Structure and Reciprocity in the Field. Management Science 62(9):2639-2650. <http://dx.doi.org/10.1287/mnsc.2015.2275>

Full terms and conditions of use: <http://pubsonline.informs.org/page/terms-and-conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

When $3 + 1 > 4$: Gift Structure and Reciprocity in the Field

Duncan S. Gilchrist

Wealthfront Inc., Palo Alto, California 94301, duncan@wealthfront.com

Michael Luca, Deepak Malhotra

Harvard Business School, Boston, Massachusetts 02163
{mluca@hbs.edu, dmalhotra@hbs.edu}

Do higher wages elicit reciprocity and lead to increased productivity? In a field experiment with 266 employees, we find that paying higher wages, per se, does not have a discernible effect on productivity in a context with no future employment opportunities. However, structuring a portion of the wage as a clear and unexpected gift—by offering an unconditional raise after the employee has accepted the contract—leads to higher productivity for the duration of the job. Gifts are roughly as efficient as hiring more workers.

Keywords: economics; behavior and behavioral decision making; microeconomic behavior; labor; IT policy and management; electronic commerce; electronic markets and auctions; organizational studies; motivation-incentives

History: Received July 1, 2014; accepted May 9, 2015, by Uri Gneezy, behavioral economics. Published online in *Articles in Advance* January 19, 2016.

1. Introduction

Economists have long recognized that employees are often paid more than the market clearing wage and that unemployed workers do not bid wages down to the point where supply equals demand. The neo-classical explanation for this phenomenon comes in the form of efficiency wage theories, which generally argue that employees will work harder when they receive high wages because they do not want to lose a high-paying job (Katz 1986). This type of mechanism relies on repeated interactions between employers and employees. In one-time jobs without any consideration for future employment, the neoclassical model would argue that efficiency wages do not increase productivity.

At the same time, a robust literature based in psychology and behavioral economics demonstrates that people care about fairness and that fairness considerations may create incentives for reciprocation. If one gives a gift to someone, that person might reciprocate, even in a one-shot game with no potential for future interaction (Gouldner 1960, Cialdini 1993, Berg et al. 1995, Fehr and Gächter 2000, Pillutla et al. 2003, Falk et al. 2008). This principle has been implemented in field settings as well. For example, Falk (2007) shows that including a small gift in a fund-raising letter leads to higher donation levels.

In the context of labor economics, fairness concerns and reciprocity have been offered as an explanation for efficiency wages (Akerlof 1982, Akerlof and Yellen 1990, Fehr et al. 2009). If employees view high

wages as a gift, then they may reciprocate by working harder, even though there is no financial incentive to do so. The thrust of this argument is that the market wage serves as a reference point, and employees will reward positive deviations from this reference point—even in a one-shot employment contract with no career concerns.

Do employees work harder when they are paid more? Laboratory experiments have mostly shown that paying an unconditional bonus before the work starts causes workers to reciprocate by working harder (e.g., Fehr et al. 1993, Fehr and Gächter 2000, Hannan et al. 2002, Charness 2004). In seminal work in the field, Gneezy and List (2006) look at two field settings, hiring roughly 40 workers to test whether paying higher-than-market wages increases output in a library data entry task and in door-to-door fund-raising. Specifically, Gneezy and List compare the output of workers assigned to a “gift” treatment (in which workers are hired at a low wage and then offered a raise immediately before starting work) with the output of workers assigned to a “nongift” treatment (in which workers are hired at and paid the low wage). They find that workers who receive the gift (i.e., the additional money) exert higher productivity for the first few hours of the task; in their setting, the temporary increase in productivity does not justify its cost.

Since then, there has been mixed evidence on the relevance of gift exchange in different contexts. Hennig-Schmidt et al. (2010), Kube et al. (2012, 2013),

Al-Ubaydli et al. (2015), and concurrent but independent work by Esteves-Sorenson and Macera (2013) all find that gifts have mild to no effects when hiring undergraduate students for field experiments. However, looking at a tree-planting firm, Bellemare and Shearer (2009) find that a one-day, unexpected gift leads to an increase in productivity. Cohn et al. (2015) also find evidence in favor of gift exchange—but only when employees view the initial wage as being unfairly low. Looking at managers in a large multinational company, Ockenfels et al. (2014) find that bonuses that fall behind natural reference points lead to lower job satisfaction and performance. There is an open question about the extent to which all of these effects persist.

However, because this prior research has not varied the base wage, the setup cannot identify whether reciprocity is triggered because the higher wages are above the market rate or because the higher wages are salient to workers who unexpectedly receive the gift after having already agreed to a job. In other words, wage amount (high versus low) is confounded with wage structure (salient gift versus no gift) in the prior literature. In this paper, we offer a large-scale gift exchange experiment in a field setting where we vary whether or not a worker receives an unexpected raise before starting work as well as the base wage offered to potential hires. This subtle difference allows us to differentiate between the impact of salient “gifts” on performance and the impact of above-market wages on performance.

Our results show that the way in which a wage is structured (and not simply its level) may be essential to generating reciprocity. In our study, hiring at and paying workers a high wage (\$4) leads to no increase in productivity; i.e., it results in an economically identical and statistically indistinguishable amount of productivity relative to hiring at and paying workers a lower wage (\$3). However, hiring workers at a lower wage and then offering them an unexpected raise ($\$3 + \1) significantly increases performance. Common models of reciprocity in labor markets (e.g., Akerlof 1982, Akerlof and Yellen 1990) assume that high wages alone are the determinant of reciprocation, but these models do not differentiate between our $3 + 1$ and 4 offers. To our knowledge, field experiments have similarly not differentiated between these two treatments.

Our experiment takes place on Upwork,¹ an online labor market with several million registered contractors. Using the Upwork platform allows us to vary wages and gifts in a setting where workers are

accustomed to tasks such as ours. Because we conduct a natural field experiment, the employees do not know that they are part of an experiment. Indeed, the Upwork platform is a powerful setting for an experiment such as ours because using it allows us to compare and control for the entire Upwork work histories of our employees. Furthermore, the Upwork marketplace allows us to conduct targeted hiring by directly inviting workers to take up our job instead of simply posting a job publicly and waiting for applications. This means we are able to hire workers at different base wages (without individuals knowing how much others have been paid) so that we can test whether it is the base wage amount or the salience of the gift that affects performance. Normally, selection would be a concern when hiring at different wages, but we are able to address the potential for selection through the combination of high take-up rates (the take-up rate in our study is 95% among workers with prior experience), comparison of characteristics across treatments for job takers (they are not significantly different), and a conservative robustness test that we will discuss later.

Our experimental design, which we describe in more detail in the next section, proceeds by hiring three groups of Upwork workers for a data entry task, all of whom have requested wages of less than \$3 per hour according to their Upwork profiles. We are clear in our recruitment messages that this is a one-time job. The first group is hired at \$3 per hour (i.e., the “3” condition). The second group is also hired at \$3 per hour, but before starting work, workers in this group are told that we unexpectedly have extra money in the budget and will pay an extra \$1 per hour, so that the total they will receive is \$4 per hour (“ $3 + 1$ ”). The third group is hired directly at \$4 per hour, so the fact that we are paying them the higher (above-market) wage does not signal a gift in a salient way (“4”). To increase the validity of the results, we choose a data entry task (to be described in more detail later) that is fairly common in online labor markets, and we only recruit workers who self-report data entry as a specialty on their Upwork profiles.

Consistent with the notion of reciprocity, we find that higher wages that include a salient gift ($3 + 1$) lead to higher and more persistent productivity across our task relative to the other two groups (3 and 4). Specifically, paying $\$3 + \1 yields a 20% increase in productivity compared with paying \$4, with no additional cost. Compared with paying \$3, paying $\$3 + \1 results in a 20% increase in productivity with a 33% increase in cost. Notably, we find that varying the base wage from \$3 to \$4 in the original contract has no statistically distinguishable effect on productivity—in fact, the point estimate of the effect is 0. To our knowledge, our experiment is the first to include the

¹ Since this experiment was completed, oDesk has become Upwork (<https://www.upwork.com/>, accessed December 7, 2015). At the time of the experiment, Upwork was called oDesk.

high base wage (4) condition, allowing us to better understand the extent to which the salience of (versus merely offering) a gift drives reciprocity.

In addition to our main results, we also find suggestive (i.e., economically large but not statistically significant) evidence that the effect is largest for those workers for whom the +1 gift may have been most salient (e.g., experienced employees, who are familiar with the typical wage structure on Upwork). These effects remain for the duration of the task.

Altogether, our results suggest that unexpected gifts ($3 + 1$) can increase productivity, even relative to offering the same above-market wages (4) at the time of hiring. One possible explanation for our results is the model of Hart and Moore (2008), which argues that contracts (in this case, for \$3 or \$4) set a strong reference point for employees and that deviations from this reference point are viewed as fair or unfair, thus triggering reciprocity. The extent to which these effects persist is an empirical question that requires further study.

The rest of our paper proceeds as follows. In §2, we outline our experimental design. In §3, we present our findings, and in §4, we conclude.

2. Experimental Design

Our experimental methodology proceeded in two steps. First, we selected a sample of Upwork workers, randomized those workers across treatments, and invited the treated workers to accept our job. Second, treated workers who responded were sent a link to a website where they could complete our data entry task. For workers in the $3 + 1$ condition, this message also provided notification of a change—a \$1-per-hour increase—in their wages. This section describes the sample, the treatment conditions, and the task itself in more detail.

2.1. Sample

We began our sample selection by restricting the universe of Upwork workers to those who claim to specialize in data entry. In particular, we required that (a) worker profiles are returned by a keyword search for “data entry” and (b) workers classify themselves as administrative support workers with a data entry specialty.² Next, we further restricted the sample to workers that list a requested hourly wage between \$2.22 and \$3.33. Since Upwork charges workers a 10% fee on gross earnings, this restriction amounts to requiring that workers request net wages between \$2 and \$3 per hour.

It is relatively easy for someone to create an Upwork profile, regardless of whether the individual

actually plans to seek work on Upwork, so in an attempt to select only serious candidates, we further restricted the sample by requiring that workers had (1) logged into Upwork within the last 30 days and (2) were listed as “Upwork ready,” which means they had passed a multiple-choice exam that tested their knowledge of the Upwork interface. Finally, to ensure workers were autonomous and not making decisions within a larger organization, we required that workers were listed as independent contractors, meaning they were unaffiliated with a company.

A total of 17,482 workers satisfied these joint criteria at the time of data collection. From this set of workers, we randomly selected 540 workers and allocated them randomly across our three treatments.³ Panel A of Figure 1 presents the initial recruitment messages with which we invited selected workers to take up our task. Of the 540 workers we invited to take up our task, 266 accepted our job offers. We exclude 36 of these workers because of a technical glitch described below, so our final sample comprises 230 workers. Notably, among workers with prior experience, our take-up rate was 95%.

2.2. Treatments

As we discussed above, we randomized workers into three treatment groups. Workers in two of the treatment groups were initially offered \$3 per hour, whereas workers in the third group were initially offered \$4 per hour. (Technically, we would pay \$3.33 and \$4.44, respectively, because 10% of gross wages go to Upwork. For simplicity, we refer to wages in net terms throughout the paper.)

We refer to the two groups of workers who had been offered \$3 per hour as the “no gift” (3) and “salient gift” ($3 + 1$) treatments, and we call the \$4 per hour treatment the “gift” condition. After accepting our job offer, workers in all treatments were reminded of the task instructions and presented with a link to a Web page where the task was located. Workers in the salient gift treatment ($3 + 1$) were additionally notified in the same message that we “have a bigger budget than expected . . . [and that] we will pay . . . \$4 per hour instead of \$3 per hour.”⁴ Panel B of Figure 1 presents the messages we sent to workers to let them know we had agreed to hire them, and Figure 2 summarizes the experimental setup.

Before proceeding, we note that although it is difficult to ascertain what constitutes the market wage for

³ We oversampled the \$3-per-hour treatment because those workers participated in a separate exercise, independent of this experiment, which took place after this experiment was completed.

⁴ This message was phrased so as not to emphasize any sacrifice the firm was making in paying a higher wage or emphasize our desire that a worker reciprocate. However, our design does not allow us to analyze the impact of the specific wording of this message.

² We note that Upwork specialties are self-reported, so neither of these two restrictions required a worker to actually have any experience in data entry on the Upwork platform (or elsewhere).

Figure 1 Job Offer Messages**Panel A: Recruitment messages***All treatments*

We are currently looking to hire a group of people to help with simple data entry. The job consists of looking at a photograph of a word and typing that word into the space provided. This is a four-hour job, with the goal of entering as much data as possible while minimizing the number of mistakes. Specifically, we need as many correctly entered words as possible in four hours because we need the data for a future task and only correct entries can be used. You will have seven days to complete the task.

You will be paid \$3 (\$4) per hour. Therefore, your total payment for the four hours will be \$12 (\$16). We hope you will accept this job.

Panel B: Acceptance messages*Treatments: 3 and 4*

Great, you are hired.

By the way, we want you to know that this is a one-time job; we do not expect to have more work in the future.

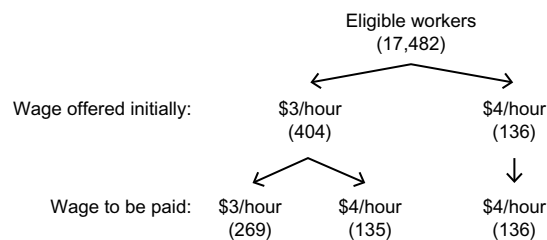
Below, you will find a link to a page where you will do the data entry. As we mentioned, the job consists of looking at a photograph of a word and typing that word into the space provided. Please enter words for four hours, after which you will be ineligible to receive further pay. Finally, please take no more than a week. We will not accept work done more than seven days after you receive this assignment.

Link to job: [here](#)*Treatment: 3 + 1*

Great, you are hired. As it turns out, we have a bigger budget than expected. Therefore, we will pay you \$4 per hour instead of \$3 per hour, so the total you can earn is \$16.

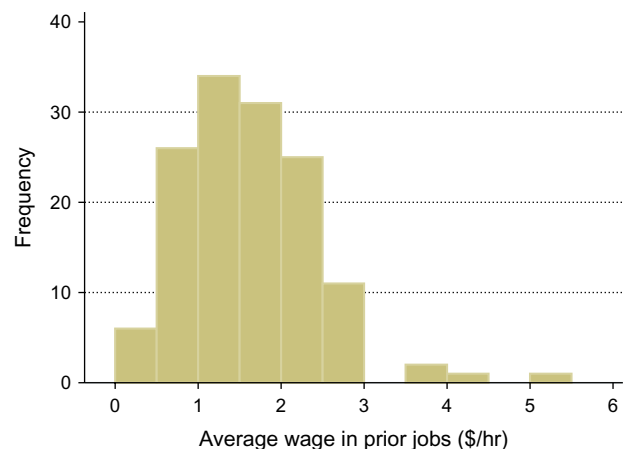
By the way, we want you to know that this is a one-time job; we do not expect to have more work in the future.

Below, you will find a link to a page where you will do the data entry. As we mentioned, the job consists of looking at a photograph of a word and typing that word into the space provided. Please enter words for four hours, after which you will be ineligible to receive further pay. Finally, please take no more than a week. We will not accept work done more than seven days after you receive this assignment.

Link to job: [here](#)**Figure 2** Experimental Design

a specific task at a particular point in time, even our lowest offer of \$3 per hour appears generous compared with wages paid to our workers in their prior roles on Upwork. Figure 3 presents a histogram of the average prior wage earned by experienced workers invited to take up our task.⁵ The graph shows that prior wages are roughly normally distributed, with a peak in the density just below the median of \$1.52 per hour. For 97% of workers, average prior wages fall below \$3 per hour. Since looking only to each worker's average prior wage might disguise variance in the underlying wages, we also compute the average percentage of prior jobs held that paid

⁵ The picture is nearly identical for the subset of workers who accept our job: as we show below, only 5% of experienced workers do not accept our offer of employment. Our focus on prior wages instead of requested wages is motivated primarily by the fact that requested wages are self-reported and thus difficult to interpret. Moreover, there is little variation in the distribution of requested wages: requested wages in our sample are essentially bimodal, with 79% of workers requesting either \$2 or \$3 per hour.

Figure 3 (Color online) Distribution of Prior Wages

Notes. This figure presents a histogram of average prior wages. The sample is all invited workers, restricted (by necessity) to those with prior experience.

at least \$3 and \$4 per hour. After weighting jobs by hours worked, on average, only 5.3% of prior jobs paid at least \$3 per hour, and only 1.7% of prior jobs paid at least \$4 per hour.⁶ We infer that our task was likely highly desirable even at our lowest wage of \$3 per hour; this is consistent with our high take-up rates (particularly among experienced workers), discussed below.

⁶ These percentages decrease slightly to 3.8% and 1.6%, respectively, if we compute the overall percentage of prior jobs held that pay at least \$3 or \$4 per hour as opposed to the within-worker statistics being averaged across workers.

Figure 4 (Color online) CAPTCHA Task


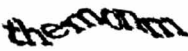

The screenshot shows a web browser window with the address bar displaying "odcaptchas.appspot.com/pcR8u/solve". The page content includes a "Welcome" message, instructions on how to solve CAPTCHAs, a list of five pointers, and a table for entering solutions. The table has two columns: "CAPTCHA" and "Solution". There are three rows of CAPTCHAs: "province" with a distorted image of "terda", "tendency" with a distorted image of "thermum", and "venture" with a distorted image of "may". Each row has an empty text box for the solution. A "Submit" button is at the bottom right.

Welcome

Solve as many as you can, leave difficult ones blank. Use the "Tab" key to move to next CAPTCHA. Use the "Enter" key to submit your solutions.

A few key pointers to keep in mind:

1. You do not get penalized for skipping a word. The only thing that matters is how many total words you enter correctly.
2. If a word is too difficult to read, you can skip it so that you don't slow down.
3. Not all of the pictures will be of real English words.
4. Spaces don't matter.
5. Capitalization does not matter.

CAPTCHA	Solution
province 	<input type="text"/>
tendency 	<input type="text"/>
venture 	<input type="text"/>

2.3. Task

As we explained in our initial recruitment messages, the task we presented to workers asked them to correctly enter as many CAPTCHAs as possible in the four hours allotted. Figure 4 presents a screenshot of the task itself, as seen by workers. CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is a system that asks a user to transcribe a word or phrase that is presented as a picture. Many online companies use CAPTCHAs to prevent automated software from easily accessing information or making decisions without a human being involved. For example, Ticketmaster® requires potential buyers to enter a CAPTCHA before purchasing tickets in order to stop a person from using a program that repeatedly buys tickets (which is something that a scalper might otherwise do). In online labor markets such as Upwork, there is a high level of demand for people to do data entry (and in fact, even specifically to enter CAPTCHAs), which means that our task would come across as a reasonably natural request. Importantly, CAPTCHA entry requires effort but has a straightforward and simple output (correctly entered CAPTCHAs), giving employees

sufficient opportunity to “repay the gift,” all factors that Englmaier and Leider (2012) and Kessler (2013) find are crucial for gift exchange to occur.

2.4. Experimental Validity

Before proceeding to our main findings, we analyze how worker characteristics are related to treatment and job take-up.⁷ Throughout, we exclude results from 12 workers who did not complete the four hours of work, as well as 24 workers in the initial wave who were able to complete more than four hours of work as a result of a technical glitch that allowed them to exceed the time limit.⁸ Thus, of the 540 workers invited to take up our task, we are left with 230, all of whom are included in our analysis.

⁷ We remind the reader that we did not stratify the sample on any characteristics.

⁸ There was no statistically distinguishable difference across the three treatment groups in the likelihood of an employee working for more or less than four hours; we exclude these workers because allowing employees to work for different lengths of time makes it harder to compare total productivity across employees. However, including results for these workers does not change our baseline results; see Table A.1 in the appendix.

Table 1 Comparing Worker Characteristics Across Treatments

	Did not take up the job			Took up the job		
	3	3 + 1	4	3	3 + 1	4
Panel A: All workers						
<i>Experienced</i>	0.021	0.048 (0.301)	0.029 (0.752)	0.691	0.69 (0.987)	0.629 (0.409)
<i>Number of prior jobs</i>	0.162	0.238 (0.881)	1 (0.0901)	8.082	7.397 (0.719)	4.548 (0.0593)
<i>Wage requested</i>	2.646	2.678 (0.675)	2.666 (0.795)	2.741	2.738 (0.972)	2.815 (0.343)
<i>N</i>	142	63	69	110	58	62
<i>N taker/N invited</i>	—	—	—	44%	48%	47%
Panel B: Experienced workers						
<i>Number of prior jobs</i>	7.667	5 (0.757)	34.50 (0.0322)	11.70	10.72 (0.703)	7.231 (0.0842)
<i>Wage requested</i>	3.053	2.777 (0.585)	2.775 (0.623)	2.761	2.750 (0.909)	2.882 (0.212)
<i>Average prior wage</i>	1.769	1.494 (0.716)	2.233 (0.586)	1.706	1.957 (0.191)	1.908 (0.308)
<i>Average prior rating</i>	4.957	4.503 (0.285)	4.615 (0.456)	4.565	4.548 (0.913)	4.495 (0.646)
<i>Worked in last 30 days</i>	0.333	0 (0.314)	1 (0.102)	0.447	0.525 (0.429)	0.410 (0.708)
<i>N</i>	3	3	2	76	40	39
<i>N taker/N invited</i>	—	—	—	96%	93%	95%

Notes. This table presents characteristics of workers invited to take up our job, split into workers that did not take up the job (first set of three columns) and workers that did take up the job (second set of three columns). Panel A analyzes characteristics for all workers, and panel B examines experienced workers only. Means are presented in the first row, and the second row presents p -values from T -tests comparing respective characteristics in a given subsample with those of the analogous 3 treatment. The sample sizes in the calculation of *average prior wage* for takers in the 3, 4, and 3 + 1 treatments are, respectively, 66, 30, and 33, because some experienced workers had only fixed-price prior jobs. Similarly, the sample sizes in the calculation of *average prior rating* for takers in the 3, 4, and 3 + 1 treatments are, respectively, 72, 38, and 37, because some experienced workers' first and only contract was ongoing at the time of the experiment.

Table 1 presents the results. In panel A, we first present a base set of statistics for all workers, and then in panel B, we present an extended set of statistics that are available only for the subsample of experienced workers (i.e., those with at least one prior job). We separately focus on experienced workers because they exhibit much higher take-up rates, provide more data for analyses, and have greater familiarity with tasks and wages on Upwork relative to inexperienced workers. Within each panel, the first set of three columns presents statistics on key characteristics (e.g., number of prior jobs), separated by treatment group, for those who did not accept our job offer. The second set of three columns in each panel presents analogous statistics for those who did accept. The statistics presented are the mean of a characteristic by subgroup and then below, in parentheses, the p -value from a T -test comparing the values of a subgroup's covariate with the same covariate of the analogous "3" subgroup. Our main empirical analysis focuses on the productivity of employees who accepted our job offers, so our objective in this analysis is to verify both that the treatment groups are balanced and that

there is no apparent selection among employees who accepted our job offers.

Although our overall take up rate is 46%, job acceptance rates and other worker characteristics are similar across treatment groups. Overall, about two-thirds of job takers are experienced, and the take-up rate is 95% among workers with prior experience and 22% among inexperienced workers.⁹ The only notable difference across the three groups of job takers (which is still not statistically significant) is that the number of prior jobs is lower in the 4 group than in the 3 + 1 and 3 groups. Probing more closely, we see that this is because an (insignificantly) larger percentage of takers in the 4 group are inexperienced and also because experienced takers in the 4 group themselves have (insignificantly) fewer prior jobs. The former could be a cause for concern, but as we show below, our

⁹ In an Upwork experiment that uses a similar pool of data entry workers to analyze the impact of reviews on future hiring, Pallais (2014) also finds that experienced workers are more likely to take up jobs. However, the divergence is less extreme in her context: she finds that 54% and 33% of experienced and inexperienced workers accept her job offers, respectively.

results are actually most pronounced for experienced workers—and are robust to only analyzing experienced workers. Moreover, it seems unlikely that the latter feature (insignificantly fewer prior jobs among experienced workers in the 4 group) is biasing our results, given that the difference is relatively small and that take-up rates among experienced workers are all close to 100%. Thus, in aggregate, the data support the validity of our experimental design.¹⁰ Finally, to further ensure that selection is not driving results, we show below that our main results are robust to a conservative test where we code all nontakers as having completed zero CAPTCHAs.

3. Results

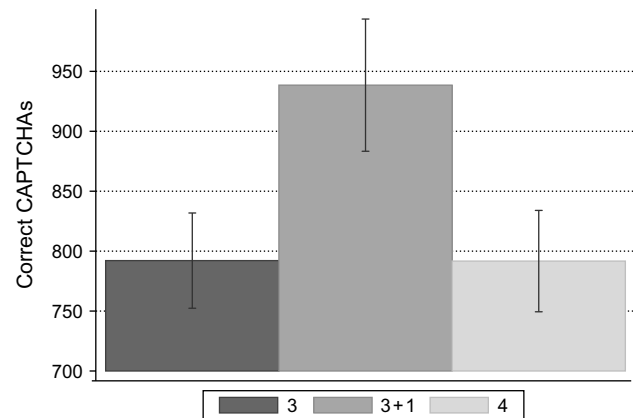
This section presents the results of our experiment. We first discuss the main effect and its persistence, and then we analyze how the effect varies with worker characteristics.

3.1. Main Effect: $3 + 1 > 4$

The gift exchange literature has posited that high wages elicit reciprocity, which could in turn rationalize above-market wages even in the context of unrepeated employment. This suggests that market wages might be a reference point and that paying more would elicit higher productivity. In this case, our 4 and 3 + 1 conditions should elicit the same response, which would be higher than the 3 condition. To our knowledge, we are the first to include something akin to our 4 treatment, which allows us to shed light on the conditions under which we should expect workers to reciprocate high wages.

Figure 5 presents the main effect graphically, and column (1) of Table 2 presents it in regression form. In particular, column (1) of Table 2 presents results from a regression of the total number of correctly completed CAPTCHAs on a constant and dummies for treatments 3 + 1 and 4, respectively (so all estimates are relative to the 3 group). The figure and the table both show that the respective numbers of CAPTCHAs correctly completed by the 3 and 4 groups are nearly identical, and we cannot reject the hypothesis that

Figure 5 The Gift Leads to Higher Productivity Than Either High or Low Base Wages



Notes. This figure shows our main result. Thick bars present means and confidence bands present one standard error relative to the mean.

they are the same. There are two potential explanations for this result. One possible explanation is that, as modeled by Hart and Moore (2008), contracts (in this case, 3 or 4) determine the reference point, which is viewed as neither fair nor unfair and hence does not elicit reciprocity. An alternative possibility is that because both wages are generously above the prior wages of employees, both elicit extra effort relative to the amount an employee would have worked at an even lower wage. In this interpretation, there is reciprocity, but it diminishes with the size of the gift, so that the fourth dollar does not increase it any further.

Our main finding is that the 3 + 1 group correctly entered 146 more CAPTCHAs relative to *both* groups over the course of the task (a 20% increase), and this result is statistically significant at the 5% level. Column (2) shows that the results are unchanged if the number of completed correct CAPTCHAs is specified in log form; column (3) shows that the treatment effect measured in the number of correct CAPTCHAs entered is not due to a change in the ratio of correct to incorrect entries. Overall, high wages elicit more effort when the wage is structured such that the gift component is made salient (e.g., by presenting it separately, as in this study). Paying \$4 elicits the same amount of productivity as paying \$3 but not the same as \$3 + \$1.

Although our sample is balanced on observables and although take-up rates were similar across all treatments (see Table 1), we conduct an additional robustness check to verify that our “ $3 + 1 > 4$ ” result is not driven by selection induced by the fact that different treatments are hired at different wages. That is, a potential explanation for the worse performance of 4 relative to 3 + 1 is that negative sorting led less able workers to refuse the job invitation for \$3 per hour

¹⁰ We note that two characteristics in the sample of “experienced workers who did not take up the job” are actually statistically significantly different across treatment groups. However, we believe this is not a cause for concern, given this subgroup’s exceptionally small sample size: each treatment in this subgroup is composed of three or fewer workers. Nevertheless, as an additional robustness check against selective take-up, we repeat the analysis in Table 1 (in which we compare characteristics across treatments) but this time pool the 3 and 3 + 1 treatments since these treatment groups were hired at the same wage. For brevity, we exclude the results here, but none of the differences between the pooled treatment groups hired at \$3 per hour and the treatment group hired at \$4 per hour are statistically significant at the 10% level.

Table 2 Performance

	(1) Correct	(2) Log correct	(3) Percentage correct	(4) Correct per quarter
<i>Wage = 3</i>	Omitted	Omitted	Omitted	Omitted
<i>Wage = 3 + 1</i>	146.4** (67.93)	0.286*** (0.100)	1.224 (1.433)	Omitted
<i>Wage = 4</i>	−0.397 (58.03)	0.0726 (0.116)	0.427 (1.384)	Omitted
<i>Quarter = 1</i>				Omitted
<i>Quarter = 2</i>				18.28*** (5.108)
<i>Quarter = 3</i>				12.15* (6.946)
<i>Quarter = 4</i>				10.12 (6.408)
<i>Wage = 3 + 1 × Quarter 1</i>				32.97** (16.23)
<i>Wage = 3 + 1 × Quarter 2</i>				30.81 (18.94)
<i>Wage = 3 + 1 × Quarter 3</i>				39.15** (19.77)
<i>Wage = 3 + 1 × Quarter 4</i>				43.49** (19.15)
<i>Wage = 4 × Quarter 1</i>				0.528 (14.33)
<i>Wage = 4 × Quarter 2</i>				−9.318 (15.32)
<i>Wage = 4 × Quarter 3</i>				−1.359 (16.14)
<i>Wage = 4 × Quarter 4</i>				9.168 (16.06)
<i>Constant</i>	792.1*** (39.82)	6.467*** (0.0830)	84.11*** (1.045)	186.9*** (9.884)
<i>N</i>	230	230	230	920
Adjusted <i>R</i> ²	0.017	0.017	−0.006	0.0014
Effect of $3 + 1$ relative to 4	146.8** (69.36)	0.214** (0.990)	0.797 (1.335)	
Effect of $3 + 1$ relative to 4 by quarter				
<i>[Wage = 3 + 1 × Quarter 1] − [Wage = 4 × Quarter 1]</i>				32.44* (16.54)
<i>[Wage = 3 + 1 × Quarter 2] − [Wage = 4 × Quarter 2]</i>				40.13** (18.56)
<i>[Wage = 3 + 1 × Quarter 3] − [Wage = 4 × Quarter 3]</i>				40.51** (20.22)
<i>[Wage = 3 + 1 × Quarter 4] − [Wage = 4 × Quarter 4]</i>				34.32* (20.50)

Notes. This table presents our main results. Column (1) analyzes the number of correct CAPTCHAs by treatment, and column (2) does the same but gives the outcome in logs instead of levels. The outcome in column (3) is the percentage of CAPTCHAs that were correct, and column (4) presents results from a panel regression where the outcome is the number of correct CAPTCHAs completed by each worker in each quarter of their work. For ease of interpretation, the bottom panel presents treatment effects of $3 + 1$ relative to 4. All standard errors are robust, and standard errors for column (4) are clustered at the worker level.

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

(into the $3 + 1$ treatment), but those same kinds of workers did accept the job invitation for \$4 per hour. Such a scenario could lead us to falsely conclude that $3 + 1 > 4$, when really selection was driving the result. To test whether negative selective take-up is driving our result, we run a regression where all nontakers are coded as having completed zero CAPTCHAs

and all takers are coded normally. If the differential yield is driving our result, and the incentive provided by 4 is actually equivalent to that provided by $3 + 1$, then 4 should perform at least as well as $3 + 1$ in this regression. This should be the case because this new specification estimates the average treatment effect conditional on job invitation, instead of the average

treatment effect conditional on job take up. The results are reported in column (2) of Appendix Table A.1. The $3 + 1$ treatment still performs better than 4, and the difference remains statistically significant at the 10% level. We conclude that our results are not driven by selection.

Aside from the fact that we observe $3 + 1 > 4$, the extent to which our salient gift treatment outperforms the no-gift treatment ($3 + 1 > 3$) contrasts with Gneezy and List (2006), Hennig-Schmidt et al. (2010), Kube et al. (2012) and concurrent but independent work by Esteves-Sorenson and Macera (2013), all of whom find that gifts have mild to no effects when hiring undergraduate students for field experiments (although Kube et al. 2012 find that nonfinancial gifts do have a significant effect). Our result is consistent with Paarsch and Shearer (2009) and Cohn et al. (2015), who find that raises in pay lead to higher effort. As we mentioned above, one possible factor contributing to this difference is that, in all of these settings, workers presumably had little or no experience with the task they were assigned. As a consequence, they likely had weak prior beliefs about the task, typical wage levels, and how wages would be presented, which may have reduced the salience—and hence, influence—of any gifts received.

Overall, these results suggest that the extent to which a gift is salient—i.e., not just “what you pay” but “how you pay”—could have important implications for the extent to which gift exchange occurs. We next examine the persistence of the main effect, and then in §3.3, we explore how the impact of the salient gift interacts with worker characteristics, which might further accentuate the salience of the “salient gift” condition.

3.2. Persistence of the Effect

The final column in Table 2 examines how the treatments impacted worker performance over the course of the four-hour task. Our database recorded the time stamp of every CAPTCHA entry, allowing us to examine the time series of responses. We gave workers one week to complete the task, and workers were not required to complete the four-hour task in a single (uninterrupted) session.¹¹ To analyze the persistence of performance, we break the time series of entries into four quarters by first truncating any gaps between CAPTCHA entry that are longer than 10 minutes to just 10 minutes and then breaking the total time between the first and the last CAPTCHA entry (with truncated breaks) into four equal blocks

of time.¹² Thus, each block of time represents one-quarter of the time a worker spent working on our task, or roughly one hour, although the sum of the four quarters is not exactly four hours because of the way we truncate large gaps between CAPTCHA entries. The dependent variable in this regression, then, is the number of correct CAPCHAs entered in each quarter by each worker (so observations are at the worker-quarter level; we cluster standard errors by worker).

The interacted treatment-quarter coefficients show that the workers in the $3 + 1$ salient gift treatment completed an average of 30–40 more correct CAPCHAs in each quarter of the task relative to the 3 group, and three of the four coefficients are significant at the 5% level. Meanwhile, output for workers in the 4 group is not statistically distinguishable from output for workers in the 3 group for any of the quarters. To examine the effect of $3 + 1$ relative to 4 over the course of the experiment, we present in the bottom part of Table 2 the difference of the treatment-quarter estimates. The salient gift treatment consistently outperforms the 4 treatment in all quarters by 30–40 CAPCHAs.

We infer from these results that the salient gift treatment increased productivity across the length of the task, i.e., four hours of work spread over as much as a week. Although our results do not speak to the longer-run persistence of the effects of a single gift, they do show that the impact of a gift is not always as ephemeral as some of the previous evidence has suggested.

3.3. Supplementary Analysis: Exploring the Role of Worker Characteristics

Our main results show that high wages increase productivity only when it is salient that the wage is high. In this section, we explore how the impact of the salient gift varies with worker characteristics. One might expect that the impact of the salient gift would be higher for workers who have more familiarity with how wages are typically structured (i.e., without gifts) in this context. For example, workers with prior Upwork experience may be more influenced than those who have not used Upwork before. Before turning to the results, we note that the sample sizes analyzed in this section are relatively small, so although the results are suggestive, the differences-in-differences across demographic cuts are not statistically significant at conventional levels.

¹¹ Upwork’s platform logged and billed employers for the hours a worker was logged into Upwork’s task-specific interface. Employers were given the option of specifying a time limit, which, in our case, was set to four hours.

¹² We chose 10 minutes as a natural length of time to use for truncation because the Upwork platform verifies its workers’ focus by taking randomly timed screenshots, and 10 minutes is the stated approximate time between screenshots. Thus, a break longer than 10 minutes is likely to represent true time away from the task.

Table 3 Performance and Worker Characteristics

Treatment	3	3 + 1	4	3	3 + 1	4
Panel A: All workers						
	<i>No experience</i>			<i>Experience</i>		
Mean	834.4	861.8	734	773.1	973.0**	825.7
Std. error	(73.74)	(76.39)	(74.58)	(47.29)	(72.13)	(50.81)
N	34	18	23	76	40	39
Panel B: Experienced workers						
	<i>Number of prior jobs < Median</i>			<i>Number of prior jobs > Median</i>		
Mean	693.9	891.0	841.2	827.7	1,055.1*	807.7
Std. error	(75.58)	(93.73)	(60.36)	(59.91)	(108.9)	(86.40)
N	31	20	21	45	20	18
	<i>Did not work in last 30 days</i>			<i>Worked in last 30 days</i>		
Mean	807.7	928.6	862.7	730.5	1013.2**	772.7
Std. error	(66.35)	(76.13)	(70.93)	(67.13)	(120.2)	(70.80)
N	42	19	23	34	21	16
	<i>Average prior wage < Median</i>			<i>Average prior wage > Median</i>		
Mean	722.9	1,111.6**	700.8	869.5	952.3	816.7
Std. error	(65.85)	(147.9)	(84.06)	(72.59)	(78.76)	(83.02)
N	35	16	12	31	17	18

Notes. This table analyzes how the number of correct CAPTCHAs varies with worker characteristics. Each set of statistics presents means and standard errors by treatment (listed at the top of the table) and subsample. Panel A presents results for all workers, whereas panel B restricts the sample to experienced workers only. The median number of prior jobs is 6 and the median prior wage is \$1.50 per hour.

** and * indicate statistical significance at the 5% and 10% levels, respectively, of the 3 + 1 and 4 treatments relative to the 3 treatment (in the same subsample).

Table 3 shows the mean number of correctly completed CAPTCHAs and its standard error, organized horizontally by treatment (3, 3 + 1, and 4) and demographic category (e.g., experienced versus inexperienced) and vertically by different demographic subgroups (e.g., different levels of experience). Panel A presents results for all workers, whereas panel B presents results for experienced workers only, and asterisks denote the significance of treatments 3 and 4 relative to the relevant 3 + 1 group. The table shows that the estimated effect is indeed most pronounced for employees with characteristics associated with strong priors: it is strongest for experienced employees and, among experienced employees, those with more prior jobs, and those who have worked most recently. The implications of these results differ from (but are not inconsistent with) the findings of Cohn et al. (2015), who find that gifts can have real effects on productivity, but they are most effective for workers who perceive the base wage as unfair. Since our wages are all well above historical wages (as discussed in §2.2), it seems unlikely that any of our workers feel as if they are being unfairly underpaid. At the same time, judgments of fairness are still relevant to our results in that a strictly selfish worker would not be expected to work harder in the 3 + 1 group.

The final set of estimates suggests that the salient gift's generosity appears to play a role as well: the estimated effect is largest for workers with the lowest prior wages (but the difference in effect size is not

statistically significant). An effect of this magnitude actually suggests that, for the workers who previously earned wages below the median of our sample, the 3 + 1 treatment is more efficient than the 3 treatment in terms of CAPTCHA completions per dollar expenditure. The \$1-per-hour gift (3 + 1) increased average CAPTCHA completion from 723 to 1,112 for this group, a 54% increase in productivity, at a cost increase of only 33%. In other words, for a subset of (targetable) employees, the 3 + 1 wage structure is more efficient than both the 3 and the 4 wage structures. (Our main results had earlier shown that 3 + 1 is efficient relative to 4.)

4. Discussion

We find that providing employees with an unexpected pay increase can increase productivity—even when there is no prospect for future employment. However, high wages that actually look like the types of efficiency wages we usually see in the field had no discernable effect. In other words, employees exhibit more reciprocity when the gift is more salient. Taken in aggregate, this suggests that *how you pay* can be as important as *what you pay*. In this section, we further discuss the managerial implications of our results, as well as limitations of our work and areas for future research.

Overall, our findings demonstrate that managers face a trade-off in deciding whether to increase output by hiring more employees or by giving gifts to existing employees. The elasticity that we measure

implies that a company that cares only about productivity would approximately break even between hiring more employees and giving gifts to a smaller set of employees. This raises the possibility that targeted gifts may be efficient in some environments and not in others.

In thinking through the design of gifts in compensation packages, managers may want to consider which types of employees are most responsive to gifts and the contexts in which gifts are most likely to be salient. To this end, Englmaier and Leider (2012) and Kessler (2013) explore factors that predict when reciprocity in labor markets is likely to occur (e.g., how much a manager benefits from worker effort, the complexity of the task).

Managers should also consider the characteristics of a gift that will make it most salient (e.g., the gift's unexpected nature, its being labeled more conspicuously, the amount of publicity it will garner). For example, our findings suggest that firms that are looking to be more generous to employees might benefit from "labeling" the high wages and other gifts that they give to employees, rather than simply assuming that employees are correctly inferring the intention behind a wage. To see how an organization might go about labeling gifts, consider the recent decision by the popular clothing store Gap® to raise its minimum wage offered to U.S. employees to a rate that exceeds local and national standards (Greenhouse 2014). Many employees are now earning a wage that is above the minimum and above that offered by neighboring stores. When an employee is hired, Gap could simply state the starting wage in an offer letter and assume that the employee knows that this is generous. Alternatively, Gap could include a few lines in the letter explaining to the employee that the offered wage is in fact higher than a competitor analysis would dictate, but that the company wants to treat its employees well and pay a fair living wage. Our findings suggest that this type of discussion could increase performance.

Our findings additionally suggest that the timing of a gift factors into gift salience. In an organizational context, firms might find that a bonus that is given independently may be more powerful than one that is included in a standard paycheck. Likewise, our findings suggest that the same gift may be more or less salient to certain types of employees based on their prior expectations: a gift that is quite unprecedented in size or type may have a greater effect on employees with greater tenure than on those who are new to the organization. Meanwhile, the impact on productivity may erode over time for gifts that show up each year at the same time (e.g., a holiday bonus).

Altogether, our results provide new evidence that "how" employers pay affects productivity. Consistent with Roth's (2002) description of the "economist

as engineer," understanding how to structure wages more effectively—in particular, beyond the teachings of the neoclassical model—presents a promising direction for continued future research.

Acknowledgments

The authors thank Kristiana Laugen for exceptional research assistance and Upwork for implementational support. The authors are also grateful to Uri Gneezy, John Horton, an associate editor, and three referees for their thoughtful comments.

Appendix

Table A.1 Robustness

	(1) Correct	(2) Correct
<i>Wage</i> = 3	Omitted	Omitted
<i>Wage</i> = 3 + 1	161.6** (64.20)	133.4** (56.98)
<i>Wage</i> = 4	4.724 (55.56)	18.16 (48.83)
<i>Constant</i>	771.3*** (37.57)	364.1*** (29.45)
<i>Sample</i>	Taker + Excluded	Taker + Excluded + Nontaker
Effect of 3 + 1 relative to 4 [<i>Wage</i> = 3 + 1] – [<i>Wage</i> = 4]	156.9** (66.22)	115.2* (62.43)
<i>N</i>	266	540
Adjusted <i>R</i> ²	0.022	0.009

Notes. This table analyzes the potential for selection in our analysis. Column (1) repeats the baseline analysis presented in the first column of Table 2 but includes results for the 36 workers who did not complete exactly four hours of work because of a technical glitch. Column (2) does the same but additionally includes the 274 workers who did not accept our job offer (these workers are coded as having completed zero CAPTCHAs). Robust standard errors are presented in parentheses.

* $p < 0.1$; ** $p < 0.05$.

References

- Akerlof GA (1982) Labor contracts as partial gift exchange. *Quart. J. Econom.* 97(4):543–569.
- Akerlof GA, Yellen JL (1990) The fair wage-effort hypothesis and unemployment. *Quart. J. Econom.* 105(2):255–283.
- Al-Ubaydli O, Andersen S, Gneezy U, List JA (2015) Carrots that look like sticks: Toward an understanding of multitasking incentive schemes. *Southern Econom. J.* 81(3):538–561.
- Bellemare C, Shearer B (2009) Gift giving and worker productivity: Evidence from a firm-level experiment. *Games Econom. Behav.* 67(1):233–244.
- Berg J, Dickhaut J, McCabe K (1995) Trust, reciprocity, and social history. *Games Econom. Behav.* 10(1):122–142.
- Charness G (2004) Attribution and reciprocity in an experimental labor market. *J. Labor Econom.* 22(3):665–688.
- Cialdini R (1993) *Influence: Science and Practice* (HarperCollins College Publishers, New York).
- Cohn A, Fehr E, Goette L (2015) Fair wages and effort provision: Combining evidence from a choice experiment and a field experiment. *Management Sci.* 61(8):1777–1794.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 19:44 . For personal use only, all rights reserved.

- Englmaier F, Leider S (2012) Managerial payoff and gift exchange in the field. Working paper, Ludwig Maximilian University, Munich, Germany.
- Esteves-Sorenson C, Macera R (2013) Revisiting gift exchange: Theoretical considerations and a field test. Working paper, Yale University, New Haven, CT.
- Falk A (2007) Gift exchange in the field. *Econometrica* 75(5): 1501–1511.
- Falk A, Fehr E, Fischbacher U (2008) Testing theories of fairness: Intentions matter. *Games Econom. Behav.* 62(1):287–303.
- Fehr E, Gächter S (2000) Fairness and retaliation: The economics of reciprocity. *J. Econom. Perspect.* 14(3):159–181.
- Fehr E, Goette L, Zehnder C (2009) A behavioral account of the labor market: The role of fairness concerns. *Annual Rev. Econom.* 1:355–384.
- Fehr E, Kirchsteiger G, Riedl A (1993) Does fairness prevent market clearing? An experimental investigation. *Quart. J. Econom.* 108(2):437–459.
- Gneezy U, List JA (2006) Putting behavioral economics to work: Testing for gift exchange in labor markets using field experiments. *Econometrica* 74(5):1365–1384.
- Gouldner A (1960) The norm of reciprocity: A preliminary statement. *Amer. Sociol. Rev.* 25(2):161–178.
- Greenhouse S (2014) Gap to raise minimum hourly pay. *New York Times* (February 19), <http://www.nytimes.com/2014/02/20/business/gap-to-raise-mini-mum-hourly-pay.html>.
- Hannan RL, Kagel J, Moser D (2002) Partial gift exchange in an experimental labor market: Impact of subject population differences, productivity differences, and effort requests on behavior. *J. Labor Econom.* 20(2):923–951.
- Hart O, Moore J (2008) Contracts as reference points. *Quart. J. Econom.* 123(1):1–48.
- Hennig-Schmidt H, Sadrieh A, Rockenbach B (2010) In search of workers' real effort reciprocity: A field and a laboratory experiment. *J. Eur. Econom. Assoc.* 8(4):817–837.
- Katz L (1986) Efficiency wage theories: A partial evaluation. Fischer S, ed. *NBER Macroeconomics Annual 1986*, Vol. 1 (MIT Press, Cambridge, MA), 235–290.
- Kessler JB (2013) When will there be gift exchange? Addressing the lab-field debate with a laboratory gift exchange experiment. Working paper, University of Pennsylvania, Philadelphia.
- Kube S, Maréchal MA, Puppe C (2012) The currency of reciprocity: Gift exchange in the workplace. *Amer. Econom. Rev.* 102(4):1644–1662.
- Kube S, Maréchal MA, Puppe C (2013) Do wage cuts damage work morale? Evidence from a natural field experiment. *J. Eur. Econom. Assoc.* 11(4):853–870.
- Ockenfels A, Sliwka D, Werner P (2014) Bonus payments and reference point violations. *Management Sci.* 61(7):1496–1513.
- Paarsch HJ, Shearer BS (2009) The response to incentives and contractual efficiency: Evidence from a field experiment. *Eur. Econom. Rev.* 53(5):481–494.
- Pallais A (2014) Inefficient hiring in entry-level labor markets. *Amer. Econom. Rev.* 104(11):3565–3599.
- Pillutla M, Malhotra D, Murnighan JK (2003) Attributions of trust and the calculus of reciprocity. *J. Experiment. Soc. Psych.* 39(5):448–455.
- Roth AE (2002) The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica* 70(4):1341–1378.