

# How natural field experiments have enhanced our understanding of unemployment

Omar Al-Ubaydli<sup>1,2,3\*</sup> and John A. List<sup>4,5</sup>

**Natural field experiments investigating key labour market phenomena such as unemployment have only been used since the early 2000s. This paper reviews the literature and draws three primary conclusions that deepen our understanding of unemployment. First, the inability to monitor workers perfectly in many occupations complicates the hiring decision in a way that contributes to unemployment. Second, the inability to determine a worker's attributes precisely at the time of hiring leads to discrimination on the basis of factors such as race, gender, age and ethnicity. This can lead to systematically high and persistent levels of unemployment for groups that face discrimination. Third, the importance of social and personal dynamics in the workplace can lead to short-term unemployment. Much of the knowledge necessary for these conclusions could only be obtained using natural field experiments due to their ability to combine randomized control with an absence of experimenter demand effects.**

Unemployment is a major problem that causes great hardship for jobseekers<sup>1</sup>, as well as political and social instability<sup>2</sup>. Economists have provided many theories of unemployment, including macroeconomic theories that emphasize government spending<sup>3</sup>, monetary policy<sup>4</sup>, exchange rate factors<sup>5</sup> and technological transformations<sup>6</sup>, as well as microeconomic theories, such as impediments to hiring and firing<sup>7</sup>, including minimum wages<sup>8</sup>, union negotiating tactics<sup>9</sup>, imperfect job searching<sup>10</sup> and non-competitive product markets<sup>11</sup>. These rich and varied theories have yielded a wide array of policy recommendations for countering unemployment<sup>12</sup>. The performance of such policies has led to economists updating their theories, most famously during the Great Depression (1929–1941), which led many to question the prevailing doctrine of allowing unemployment to correct itself<sup>13</sup>.

Notwithstanding, the natural scientific interplay between theory and data has been impeded by the relative difficulty of gathering easily interpretable data<sup>14</sup>. Observational data suffer from the problem of correlation not implying causation, while the event-study remedy (known as the difference-in-difference method in the economics literature) is available only in sporadic and extraordinary cases, such as the Great Depression<sup>15</sup>. Alternatively, laboratory experiments are regarded as yielding insights that are not generalizable to the entire economy, because they are conducted in artificial environments with small numbers of subjects<sup>16</sup>.

Natural field experiments, which combine the realism of observational data with the randomized control of laboratory experiments<sup>17</sup>, are rarely feasible when studying economy-level phenomena such as unemployment, because decisions such as income tax levels are too important for academically minded experimentation. This has retarded the advancement of traditional unemployment theory. Yet, the past 40 years have seen the formulation of new theories suitable for testing using natural field experiments<sup>18</sup>, enhancing of our understanding of unemployment.

There exist many surveys of natural field experiments in the literature, including several surveys of labour market experiments<sup>19</sup>. This paper extends those surveys by focusing on what natural field experiments have taught us about unemployment — an issue that remains unconsidered in the literature to the best of our knowledge.

The paper has three main conclusions. First, the inability to monitor workers effectively in many occupations complicates the hiring decision in a manner that can potentially contribute to unemployment. Second, the inability to determine precisely a worker's attributes at the time of hiring leads to employment discrimination on the basis of factors such as race and gender. Third, the importance of social and personal dynamics in the workplace can create short-term frictions that lead to unemployment.

This Review focuses on exposing pivotal contributions, and those that bear the most pedagogical value, rather than providing a comprehensive review. The aim will be to shed light on the improvements in our knowledge that were made possible by the methodological advantages of natural field experiments.

## Natural field versus conventional laboratory experiments

Natural scientists frequently use experimental data to conduct their empirical investigations, as randomized control allows them to attribute variation in outcome variables of interest to the explanatory variables that they are manipulating, and to therefore test their theories incisively<sup>20</sup>. By virtue of studying humans, social scientists are typically less able to use experimental data, as experimenting on human subjects raises a series of often insurmountable ethical and feasibility issues<sup>21</sup>. As a result, social scientists will often test their theories using observational data, which makes the inference process more difficult, as attributing causality is much more challenging when explanatory variables cannot be manipulated according to the principle of randomized control; see Harrison and List<sup>17</sup> for elucidation on this distinction.

Our focus in this section is to briefly explore the difference between laboratory and natural field experiments, which natural scientists are less likely to be familiar with; for more details, see the symposium in Fréchette and Schotter<sup>22</sup>.

Conventional laboratory experiments in the social sciences have: (1) a subject pool that is college students; (2) a highly restrictive decision-making environment; (3) low stakes; and (4) a high degree of scrutiny, including informed consent.

An example of a conventional laboratory experiment follows: researchers giving college students \$10 each to participate in a study

<sup>1</sup>Bahrain Center for Strategic, International and Energy Studies, Manama, Bahrain. <sup>2</sup>George Mason University, Fairfax, VA, USA. <sup>3</sup>King Fahad University of Petroleum and Minerals, Dhahran, Saudi Arabia. <sup>4</sup>University of Chicago, Chicago, IL, USA. <sup>5</sup>NBER, Cambridge, MA, USA. \*e-mail: [omar@omar.ec](mailto:omar@omar.ec)

that has experimentally manipulated information about different charities and asking them to decide how much of their \$10 they wish to donate to the charities<sup>23</sup>.

Natural field experiments have: (1) natural subject pools; (2) natural choices available to subjects; (3) natural stakes; and (4) natural scrutiny levels, including an absence of informed consent (subjects are unaware of their being in an experiment).

An example of a natural field experiment follows: Facebook manipulated newsfeed information (without explicitly informing its users) and then researchers studied the effect of this manipulation on users' activities<sup>24</sup>.

Compared with natural field experiments, laboratory experiments have two key advantages<sup>25</sup>. First, the enhanced levels of control over the environment permit manipulations that are unfeasible in certain field settings. For example, a laboratory experiment would be the only option for experimentally studying central bank interest-rate decisions. Second, the homogeneity of the subject pool (college students) and the environment (a computer laboratory) make it comparatively easy to conduct epistemologically valuable replications<sup>26</sup>.

Natural field experiments have a key advantage over laboratory experiments for investigating certain questions. By ensuring a natural environment, including the absence of Hawthorne effects<sup>27</sup>, they deliver insights that are more likely to be relevant to the real world, known as higher levels of generalizability<sup>16</sup>, and they use a subject pool that does not have a chance to select into and out of the experimental environment<sup>17</sup>. In the context of unemployment, this is especially important, as some of the most important theories of unemployment require natural environments for them to be investigated in a more thorough manner.

### Antecedents of modern theories of unemployment

Testing a theory using a natural field experiment requires researchers' ability to observe the theory's key variables and manipulate the central explanatory variables of interest. Though the classical theory of unemployment was developed in the early twentieth century<sup>28</sup>, researchers were unable to satisfy the two conditions for natural field experiments for most of the twentieth century. Up until the 1970s, there were three main problems.

First, testing the primary, demand and supply model of labour markets<sup>29,30</sup> required researchers to be able to observe the volume of labour that firms were willing and able to procure at different wage rates — a variable that is too dynamic and erratic to measure reliably<sup>31,32</sup>.

Second, the alternative models, especially the Keynesian model<sup>13</sup>, were formulated at the level of the economy, precluding the experimental manipulation of the key explanatory variables of interest, such as the level of government spending and taxation. This in turn was partially the result of economists lacking the intellectual tools to analyse the subtle strategic interactions between firms and workers, which govern the microeconomic models that mediate the economy-level effects of variations in government spending and taxation.

Third, economists knew little about experiments<sup>33</sup>, or questioned their epistemological value in the case of laboratory experiments<sup>27</sup>.

The absence of experimentation meant an inability to make decisive inductive arguments. This was very costly to policymakers because the competing models — classical and Keynesian — made wildly divergent prescriptions about how to tackle unemployment, especially regarding the nature and degree of government intervention.

By the 1990s and 2000s, three developments<sup>34–36</sup> led to the preconditions for natural field experiments being satisfied, paving the way for the use of natural field experiments to study unemployment.

First, a sequence of economic crises during the twentieth century, most notably the Great Depression 1929–1941 (ref. 17) and

the recessions of the 1970s (ref. 37), exposed deficiencies in the prevailing theories, necessitating new ones. Second, the experimental method was developed, and was deployed in testing the previously untestable model of demand and supply<sup>35,38</sup>, which was relevant to labour markets. Third, the advent of game theory<sup>36</sup> provided the tools necessary to analyse important aspects of labour markets at a scale smaller than that of the economy, such as wage negotiations and union activity<sup>39</sup>. This marked the start of the microfoundations revolution in economics, whereby phenomena that were traditionally considered macroeconomic, including unemployment, were now analysed using microeconomic models.

We now turn to a description of the key theories developed, and the related field-experimental evidence. Note that we restrict our exposition to the theories that have been explored by natural field experiments, which is a small minority. The vast majority of models of unemployment are yet to be investigated using natural experiments for the aforementioned feasibility reasons — key variables are unobservable, or their manipulation is unfeasible.

### Asymmetric information and unemployment

One class of new models analysed information asymmetries between employers and workers, meaning that employers and workers each possess information that the other side seeks but does not have access to. The two main models are the moral hazard model<sup>40</sup>, where one of the parties conceals information about their actions, and the adverse selection model<sup>41</sup>, where one of the parties conceals information about their characteristics.

**The moral hazard model.** Employers frequently imperfectly observe the actions of workers<sup>42</sup>. Sometimes, it is due to the fact that certain dimensions are physically unobservable to the employer<sup>43</sup>: did the paramedic maintain correct hand hygiene throughout the casualty-handling process? Other times, it is because certain dimensions are too subjective: did the paramedic comfort the patient psychologically?

In both cases, employers have access to potentially helpful monitoring technologies<sup>44</sup> (security cameras, unannounced inspections). The employer can then threaten the employee with punishment in the event that unsatisfactory performance is detected, including the possibility of dismissal.

According to the worker-shirking model<sup>40</sup>, when unemployment is low, dismissed workers can easily find a new job, which makes them less fearful of losing their job. The model posits that employers who imperfectly monitor workers strategically elevate the wage and decrease hiring, leading to higher unemployment. This raises the cost of being caught shirking, since the foregone earnings from dismissal are higher, and the duration of unemployment is longer.

**The adverse selection model.** Employers frequently imperfectly observe workers' job-relevant characteristics at the time of hiring, and hiring and firing are costly<sup>45</sup>. This includes characteristics such as workers' diligence, or their ability to work in teams.

In all markets, including labour, hidden information leads to the failure to realize mutually beneficial trades<sup>41</sup>. For example, Alex is selling a used car to Jo, because Alex does not need it anymore, whereas Jo really needs it. However, Jo suspects that the car may have some hidden faults, and so insists on paying a low price for the car. Alex knows that it has no hidden faults, and therefore insists on charging a high price for the car. Alex's inability to credibly convey the absence of hidden faults leads to the trade not happening, whereas it would occur if there were no hidden information. This is known as market unravelling.

Similar reasoning can be used to understand unemployment in labour markets: employers are reluctant to hire people with hidden characteristics lest they be saddled with unproductive workers, leading to a breakdown of wage negotiations<sup>46</sup>.

The generic unemployment resulting from labour markets unravelling inevitably morphs into group-specific unemployment as employers use observable traits of prospective workers, such as gender, race, unemployment duration, religious affiliation and so on, to make inferences about hidden traits<sup>47</sup>. This constitutes labour market discrimination, and represents a class known as statistical discrimination, because it is born out of a desire to use observable characteristics as tools for overcoming imperfect information about job-relevant traits<sup>46</sup>.

On the other hand, taste-based discrimination refers to a distaste for interacting with people with certain observable characteristics, such as a white person not wanting to work in the same organization as a black person out of bigotry/animus<sup>48</sup>. It also includes the possibility of a racially blind business that behaves as though it has taste-based discrimination because its customers exhibit actual taste-based discrimination. For example, if a restaurant's patrons dislike grey-haired waiters, this creates an incentive for management to exhibit taste-based discrimination even if they are themselves non-discriminatory.

Taste-based discrimination can also lead to group-specific unemployment<sup>48</sup>: a member of the minority group cannot secure a job because employers are unwilling to tolerate their so-called undesirable traits.

As we will discuss below in 'Adverse selection', natural field experiments sometimes offer the only avenue for assessing whether the discrimination in a certain setting is statistical or taste based. The benefits of being able to distinguish extend beyond the academic, intellectual domain, into the policy arena, because the counter measures for each type of discrimination are distinct<sup>19</sup>. For example, if disabled people are less likely to be offered jobs because of the perceived costliness to the employer of making suitable workplace arrangements, then the government might consider requiring such arrangements in all workplaces or subsidizing them. In contrast, if it is due to antipathy that employers feel towards the disabled workers, despite their being equally productive, then affirmative action or sensitivity training are likely to be more effective interventions.

In both classes of discrimination, the elevated unemployment experienced by discriminated-against groups may be offset by diminished unemployment by the groups that experience favourable discrimination. Therefore, discrimination should be seen as a source of group-specific unemployment — which merits study in isolation — rather than necessarily as a source of aggregate unemployment.

**Critical insights from natural field experiments.** Here we review insights gained from the moral hazard and adverse selection models.

*Moral hazard.* Testing the moral hazard model using a natural field experiment is very difficult because the model only makes a prediction when comparing two discrete extremes: perfect versus imperfect monitoring. It does not make predictions about how unemployment changes as monitoring incrementally improves or deteriorates. Moreover, the ability of workers to be employed across a range of industries makes it virtually impossible to isolate unemployment in a given industry. Therefore, to the best of our knowledge, the closest that economists have come to testing the models using field experiments is to find evidence in support of subcomponents of the model<sup>49,50</sup>. These focus on showing that workers respond to incentives relating to worker shirking.

Nagin et al.<sup>49</sup> conducted a natural field experiment on workers in a telephone solicitation firm. Successful solicitations resulted in higher income for workers, but the solicitations were initially self-reported, with only imperfect follow-up and verification. This created the opportunity for shirking — fabricating successful solicitations. The authors found that decreasing the level of monitoring significantly increased strategic shirking, corresponding to false reports of successful solicitations. They also found significant

heterogeneity, with a large proportion of workers exhibiting insensitivity to the change in monitoring.

Boly<sup>50</sup> conducted a natural field experiment where he hired workers to grade exams. Workers were told that a random subsample of the papers they graded would be checked, and grading errors would lead to a financial penalty. When the level of monitoring was increased (a larger subsample was checked), the level of errors fell, consistent with the shirking model.

The original worker-shirking model treated output as being unobservable outside of spot checks. The refined version involves multiple dimensions of output where some are unobservable (quality) while others are observable (quantity). Shirking then appears in the unobservable dimensions. In such a situation, an indirect test of the responsiveness of shirking to incentives is to vary the reward attached to the observable dimension, and to see whether shirking increases in the unobservable dimension (unobservable to the employer; though observable to the researcher as a feature of the experimental design), due to the substitutability of effort across the different dimensions<sup>51</sup>.

The typical natural experiment involves a firm paying a random subset of its employees according to a piece rate, with the control group receiving a fixed hourly wage. To avoid gift-exchange confounds (see 'The gift-exchange model'), the piece rate is set so that the worker exerting the same effort as a typical fixed hourly wage worker would earn the same as a fixed hourly wage worker. Several studies found that piece rates increased output, as expected, but that there was no significant deterioration in quality<sup>52–54</sup>. To the best of our knowledge, the only natural field experiment that has found the classic Holmstrom and Milgrom<sup>51</sup> result is Hong et al.<sup>55</sup>, which varied the incentives associated with the different dimensions of output.

The most likely explanation for the counterintuitive finding is that since precise quality data are available to the researcher, workers probably assume that it is available to the manager (who might be the same person), meaning that excessive, strategic shirking is unwise.

Al-Ubaydli et al.<sup>56</sup> explored an additional reason, which is that the act of choosing a specific incentive scheme might convey hidden information about the manager's capacity to monitor<sup>57</sup>. For example, workers might surmise that a manager would only switch to piece-rate compensation if they have effective quality control. This latter possibility points to a weakness in natural field experiments, which is the relative inability to manipulate precisely subject expectations: the complexity of the environment means that researchers cannot simply assume that whatever statements employers make about incentives are taken at face value by workers; there is a chance that workers rationally question the veracity of the information or modify their expectations about other critical variables. This is why researchers in the field must track behaviour for long periods of time, if possible, to ensure that subject expectations converge on their long-run, natural level.

Taken together, these studies constitute only weak evidence in favour of the shirking model, though there is no strong evidence against the model, either. Ultimately, the model is not well suited for the kind of data that researchers can find at present. We can be confident that workers shirk, and that they respond to incentives, but we can say very little about whether this contributes to unemployment. In principle, conducting natural field experiments in industries undergoing fundamental changes in monitoring technology, and combining the resultant data with survey evidence garnered from in-depth interviews with employers, could contribute to substantive advances in our understanding.

*Adverse selection.* The adverse selection/discrimination model is also difficult to test using all data types: many classes of discrimination are both illegal and immoral, meaning that they will be undertaken covertly. This requires researchers to be more creative, crafting indirect ways of demonstrating the existence of discrimination.

One of the most broadly applied techniques is the audit study<sup>19,58</sup>: a researcher finds two people who are identical in as many observable dimensions as possible except the variable of interest, such as race or gender. Subjects, who are typically confederates, then engage in an activity studied by economists, and outcomes are compared. This process is repeated for a large number of pairs to increase inferential power, with the goal of attributing differences in outcomes to differences in the discriminating variable. This simple design suffers from several drawbacks.

First, such data are not experimental, since the explanatory variable is not assigned by the experimenter: with some exceptions (described below in the discussion of Bertrand and Mullainathan<sup>59</sup>), subjects cannot have their race or gender randomly manipulated by the experimenter, which makes it more difficult to interpret the findings. The difficulty is compounded by the possibility that there are variables that are causally related to the subject's type (race, gender and so on), which are observable to employers, but unobservable to the researcher.

Second, without suitable additional treatments (see our discussion of Gneezy et al.<sup>60</sup> below), the design does not permit researchers to distinguish between statistical and taste-based discrimination.

Despite these drawbacks, natural field experiments remain an important source of information on discrimination in labour markets. To provide a glimpse of the work, we highlight a handful of contributions, but point the interested reader to List and Rasul<sup>19</sup> and Bertrand and Duflö<sup>58</sup> for more complete overviews.

Neumark et al.<sup>61</sup> and Pager<sup>62</sup> used an audit design to examine race and gender discrimination. Neumark et al.<sup>61</sup> found that there is substantial gender discrimination for waiters (against men in low-price restaurants, and women in high-price restaurants), and Pager<sup>62</sup> found substantial racial discrimination against black people in a variety of entry-level jobs.

Concerns about the non-blindness of the confederate interviewees, and the possibility of subtle differences within pairs that are unobserved by the researcher, has led to the use of written audit studies as an alternative to studies such as Neumark et al.<sup>61</sup> and Pager<sup>62</sup>. These involve written applications (cover letters, curriculum vitae) from fictitious applicants, eliminating the blindness problem, permitting virtually complete control over the observable variables, and allowing for genuine experimental variation in the explanatory variable of interest.

In this spirit, Bertrand and Mullainathan<sup>59</sup> sent curriculum vitae to entry-level jobs, while randomly varying the applicant's name to convey race (people called 'Jamal' are statistically almost certain to be black, while people called 'Emily' are statistically almost certain to be white). The authors find 50% more callbacks for white-sounding names.

Complementing this work, Kroft et al.<sup>63</sup> used a written audit technique to investigate discrimination against job applicants who have been unemployed for some time. In this context, statistical discrimination may reflect the possibility that the long-term unemployed are deficient in a relevant characteristic, or that being unemployed for an extended period of time leads to a loss of productive skills (known as duration dependence). They find an almost 50% decrease in the probability of getting a callback for someone who has been unemployed for eight months compared with someone recently unemployed. The authors in these studies have difficulty distinguishing between statistical and taste-based discrimination. Yet, in Kroft et al.<sup>63</sup>, discrimination is higher in tighter labour markets, suggesting that statistical discrimination is the main force, since being unemployed for a considerable period of time when jobs are plentiful reflects more poorly on an individual's profile than unemployment does when jobs are scarce.

Numerous other studies deploy analogous techniques to investigate the effect of gender, ethnicity and other variables of interest, finding widespread evidence of discrimination, albeit varying in size (see Bertrand and Duflö<sup>58</sup> for a full account). One notable

weakness of written audit studies compared with personal ones is that they are suited primarily to entry-level jobs, and their outcome variable is callbacks, rather than actual employment and/or wages.

Beyond this, a drawback that is common to all generic audit studies is that the basic design does not permit distinguishing between taste-based and statistical discrimination. Following the natural field experimental approach of List<sup>64</sup>, Gneezy et al.<sup>60</sup> is an example of how to refine a basic audit study. The setting is not a labour market, but they examine a market transaction that is likely informative about labour issues: a personal audit study where able-bodied and disabled confederates seek in-person quotations for a car repair.

Gneezy et al.<sup>60</sup> found that disabled clients are quoted a fee that is 18% higher than their able-bodied peers, consistent with discrimination. The authors speculate that if it is statistical discrimination, it likely reflects the relative immobility of the disabled, which undermines their capacity to secure multiple, competing quotes. They conduct additional treatments where the confederate declares to the mechanic that they are seeking multiple quotes. This eliminates the penalty for being disabled, suggesting that the discrimination in the original design is statistical rather than taste based.

The refinements of Gneezy et al.<sup>60</sup> were tailored to a specific theory of statistical discrimination. There is no template that can be deployed across a wide array of contexts; researchers must be creative and deploy situational knowledge. The potential for such refinements is a key advantage that natural field experiments possess over observational data.

In summary, the evidence in favour of the adverse selection model as a source of unemployment is both strong and much stronger than that of the moral hazard model, due to the fact that the model is itself much more amenable to testing. This also means that further research in the adverse selection model is more likely to yield substantive advances in our knowledge than in the moral hazard domain, where drawing any conclusion (positive or negative) based on empirical evidence is difficult. At present, we can conclude with some confidence that employer discrimination leads to higher rates of unemployment for various groups.

### Social preferences and unemployment

Economists traditionally analyse labour markets assuming that actors maximize their financial well-being, known as egotistical materialism<sup>65</sup>. Building on the work of Keynes (originally published in 1936)<sup>13</sup>, and using the tools of game theory, the so-called gift-exchange model explores the consequence of employers and workers caring about the well-being of others, known as social preferences. In particular, when the tone of interpersonal relations in labour markets becomes important, wages no longer simply balance the demand and supply of labour; they also have an impact on the emotional side of labour relations, which can in turn have consequences for productivity and employment.

**The gift-exchange model.** The starting point in the gift-exchange model<sup>18</sup> is the worker-shirking model. Assuming egotistic materialism, workers will shirk: when the manager is away, workers might be inattentive to customers. Gift exchange assumes that some workers have social preferences, namely reciprocity: the desire to reward good deeds and punish bad ones for intrinsic reasons.

How does this relate to unemployment? The key implication of the model is that an employer seeking to elicit superior worker output should pay above-market wages<sup>18</sup>. This has two, mutually reinforcing effects. First, paying a higher wage is itself a good deed. Second, if other employers follow suit, then the market wage will rise, decreasing hiring and creating unemployment. This makes it harder for workers to secure a new job if they are laid off, which makes being given a job a good deed, too.

The threat of negative reciprocity also creates unemployment in the gift-exchange model. Workers perceive wage cuts as bad deeds;



fearing negative reprisals on the unobservable dimensions of output, financially distressed companies may prefer layoffs to wage cuts, since the retained workers usually prefer the former to the latter. Similarly, workers' sense of solidarity<sup>66</sup> and egalitarianism<sup>67</sup> will make them respond negatively to an employer who lays workers off only to hire replacements at a lower wage.

#### Natural field experimental evidence on the gift-exchange model.

The testable prediction of the gift-exchange model is that when receiving an unconditional act of kindness from an employer, the worker should reciprocate with higher output. In this manner, scholars focus on whether wage increases causally increase worker effort<sup>68,69</sup>.

Gift-exchange natural field experiments usually adhere to a list of standard criteria<sup>68,69</sup>. First, the job involves hiring workers for a one-time task, and in a class of work where there is minimal value to the worker of securing a future recommendation from the employer. This is to minimize the role of reputational concerns in explaining the worker's response to higher compensation: in the gift-exchange model, higher effort must be the result of an intrinsic desire to reciprocate a good deed, rather than an extrinsic desire to secure future good deeds as part of a long-term relationship. Therefore, researchers usually use an unskilled job, such as entry-level administrative work, and the worker pool is students or temporary employees<sup>69</sup>.

Second, if the focus is effort, which the worker can conceal from the employer, the experimenter may covertly engineer a situation whereby they can accurately measure effort while the workers still think it is unobservable<sup>68</sup>.

Third, after the workers have been hired and started working, the experimenter arbitrarily increases or decreases wages substantially for a random subsample (the treatment group), notifying the workers, possibly with an explanation; while the remaining workers (the control group) continue with the same wage. The experimenter then gauges the effect on performance<sup>69,70</sup>.

One early example from psychology is Pritchard et al.<sup>70</sup>. The task was a temporary job looking up prices in a mail-order catalogue. The authors found that arbitrarily increasing the wage led to an increase in a worker's productivity, and vice versa, but the effects were statistically insignificant.

An early study in the economics literature is due to Gneezy and List<sup>69</sup>, who examined two jobs: cataloguing books in a library for six hours, and door-to-door solicitations for a charity for two days. The worker pool was students recruited on campus. The results from the two tasks are similar to Pritchard et al.<sup>70</sup>: there is a positive but insignificant long-run effect (post-lunch), though they do find a statistically significant short-run increase in performance (over 50%) following the wage increase (pre-lunch). They attribute this to the well-established distinction in psychology between 'hot' (reflexive) and 'cold' (contemplative) decision-making<sup>71</sup>.

Subsequent work has reported mixed results. Using non-experimental data, Lee and Rupp<sup>72</sup> studied the consequences of a wage cut for airline pilots: after a week of increased flight delays, representing negative reciprocity by the pilots, performance returned to its long-run level. This is a direct analogue to the results in Gneezy and List<sup>69</sup>. Likewise, Kube et al.<sup>73</sup>, who also looked at students hired to do library cataloguing, and Al-Ubaydli et al.<sup>56</sup>, who looked at temporary employees hired to pack envelopes, used natural field experiments and found a similarly positive but statistically insignificant effect over the long run (around 10%), though neither report the Gneezy and List<sup>69</sup> hot/cold dichotomy. Yet, when Kube et al.<sup>73</sup> added a negative treatment in the spirit of Lee and Rupp<sup>72</sup>, they did find a large (27%) and statistically significant adverse effect on productivity.

Bellemare and Shearer<sup>74</sup> modified the basic design by using a tree-planting firm with long-term employees. They found that workers in the gift treatment increased productivity by around 10%, with the effect being largest for the workers with the longest tenure in the company. As explained above, interpreting such results

is more difficult because the role of social preferences cannot be isolated from reputational concerns.

Finally, List<sup>68</sup> tested gift exchange in a non-labour natural field experiment. Buyers in a sports paraphernalia exhibition randomly offer a low or high price for a baseball card, where the seller can covertly modify the card's quality at the time of purchase. According to the gift-exchange hypothesis, sellers should reciprocate high bids by delivering high quality. List<sup>68</sup> found that when he focused on non-local sellers, for whom there are no reputational concerns, there was no gift effect on quality. The same was true of local sellers before the launch of a third-party quality verification service (a company that employs experts to rate the quality of sports paraphernalia, in a manner analogous to art experts at an auction house). However, when local sellers were approached after the launch of a third-party quality verification service, the presence of reputational concerns yielded a gift-exchange effect, consistent with Bellemare and Shearer<sup>74</sup>.

Overall the field experimental evidence in favour of the gift-exchange hypothesis is weak<sup>75</sup>, though it seems to be stronger in the case of wage cuts rather than increases. However, low power, stemming from low sample sizes, is a serious issue in almost all of the cited papers<sup>69,73,74</sup>, and this is a result of the elevated expense of conducting natural field experiments. Moreover, as with many natural field experiments, there seems to be evidence of context-specificity in the findings, which is unsurprising given the complexity of the real world. More definitive conclusions require collecting more data from a wider range of contexts, such as more countries and more job types.

The context-specificity issue merits reinforcement due to the homogeneity exhibited by laboratory experiments on gift exchange. As the survey of Fehr et al.<sup>76</sup> demonstrates, the initial laboratory findings on gift exchange by Fehr et al.<sup>77</sup> — a strong and positive effect of a gift on effort — have been extremely robust. In fact, when List<sup>68</sup> conducted a laboratory gift-exchange game with the same subjects as those in the sports paraphernalia natural field experiment, and with nearly isomorphic strategies and payoffs, he found strong support for the gift-exchange hypothesis, entirely consistent with Fehr et al.<sup>77</sup>.

What explains this apparent dichotomy between laboratory and natural field results? We subscribe to the arguments presented in Levitt and List<sup>27</sup>, summarized above, and expanded on in Al-Ubaydli and List<sup>78</sup>, namely that the artificially high levels of scrutiny in laboratory experiments — specifically the subjects' belief that the researcher can see all actions — lead to an exaggeration of social preferences. In the field, despite the fact that interactions are face-to-face, certain dimensions of behaviour are observable only to one party (according to that party's beliefs).

Returning to our original question, what have natural field experiments on gift exchange taught us about unemployment? At present, the limited data suggest that gift-exchange effects may well exist, but there are doubts over their persistence a few days following a wage cut or boost. We conclude that, at best, it is likely that social preferences contribute to short-run unemployment, exacerbating the mechanisms described in the classical model of unemployment. But, in the long-run, persistent unemployment can be attributed to an alternative to the gift-exchange model.

#### Conclusions

Natural field experiments have contributed to our understanding of unemployment in three main ways. First, the inability to monitor workers perfectly in many occupations complicates the hiring decision in a way that contributes to unemployment. Second, the inability to determine a worker's attributes precisely at the time of hiring leads to discrimination on the basis of factors such as race, gender, age and ethnicity. This can lead to systematically high and persistent levels of unemployment for groups that face discrimination. Third, the importance of social and personal dynamics in the workplace can lead to short-term unemployment.

These contributions are driven by the combination of randomized control and reasonable levels of generalizability that natural field experiments afford researchers in certain contexts. In particular, extracting sound inferences on phenomena such as labour market discrimination and workplace ethics using observational or laboratory data is incredibly challenging, enhancing the value of natural field experiments.

Our finding that asymmetric information impedes hiring<sup>19</sup> has consequences for policymakers seeking to pass regulations that protect workers by imposing costs on employers, such as minimum wages<sup>79</sup> and mandated severance packages<sup>80</sup>. In particular, asymmetric information accentuates the cost of hiring in the presence of worker protections, which could indicate the need for fine-tuning and complementary policies. For example, teenagers lack labour market experience, making hiring riskier for employers. This is why several countries grant exemptions on minimum wages for young workers<sup>81</sup>.

In the context of discrimination, natural field experiments can contribute to the formulation of countermeasures. For example, Kroft et al.<sup>63</sup> study of discrimination against the unemployed was conducted at a time when US president Barack Obama was considering passing a law that banned discrimination against the unemployed<sup>82</sup>. Follow-up work that digs deeper on the nature of the discrimination (taste based versus statistical) can help indicate whether federal dollars are better spent educating the unemployed on how to legally protect themselves from such discrimination, or on how to improve their vitae in the domains that employers see as weaknesses in the unemployed.

Some government policies, particularly monetary policy, are implicitly predicated on certain beliefs regarding the effects of wage cuts on the workplace<sup>83</sup>. While natural field experiments have not done enough to change these views significantly, the tentative finding that workers get over wage cuts quite quickly<sup>52</sup> should encourage further research to assist policymakers in refining their policies.

Looking forward, scholars should consider the deployment of economy-level experimentation in labour market policies<sup>84</sup>, as part of an effort to improve our knowledge of unemployment. Policymakers might consider randomizing policies at the regional and subregional level to gain more rigorous insights as to what works, while working with institutional review boards to ensure the highest ethical standards. Ultimately, though we know much more about unemployment than economists in the twentieth century, we still have much more to learn. We trust that natural field experiments will be a key tool used to deepen our scientific understanding.

Received: 19 March 2018; Accepted: 16 November 2018;

Published online: 7 January 2019

## References

- Blakely, T. A., Collings, S. C. & Atkinson, J. Unemployment and suicide. Evidence for a causal association? *J. Epidemiol. Community Health* **57**, 594–600 (2003).
- Remmer, K. L. The political impact of economic crisis in Latin America in the 1980s. *Am. Polit. Sci. Rev.* **85**, 777–800 (1991).
- Leeper, E. M., Walker, T. B. & Yang, S. C. S. Government investment and fiscal stimulus. *J. Monet. Econ.* **57**, 1000–1012 (2010).
- Staiger, D., Stock, J. H. & Watson, M. W. The NAIRU, unemployment and monetary policy. *J. Econ. Perspect.* **11**, 33–49 (1997).
- Dornbusch, R. Expectations and exchange rate dynamics. *J. Polit. Econ.* **84**, 1161–1176 (1976).
- Long, J. B. Jr & Plosser, C. I. Real business cycles. *J. Polit. Econ.* **91**, 39–69 (1983).
- Nickell, S. Unemployment and labor market rigidities: Europe versus North America. *J. Econ. Perspect.* **11**, 55–74 (1997).
- Brown, C., Gilroy, C. & Kohen, A. The effect of the minimum wage on employment and unemployment. *J. Econ. Lit.* **20**, 487–528 (1982).
- Oswald, A. J. The economic theory of trade unions: an introductory survey. *Scand. J. Econ.* **87**, 160–193 (1985).
- Rogerson, R., Shimer, R. & Wright, R. Search-theoretic models of the labor market: a survey. *J. Econ. Lit.* **43**, 959–988 (2005).
- Benassy, J. P. Imperfect competition, unemployment and policy. *Eur. Econ. Rev.* **31**, 417–426 (1987).
- Crépon, B. & Van den Berg, G. J. Active labor market policies. *Annu. Rev. Econ.* **8**, 521–546 (2016).
- Keynes, J. M. *General Theory of Employment, Interest and Money* (Atlantic Publishers & Distributors, New Delhi, 2016).
- Engle, R. F., Hendry, D. F. & Richard, J. F. Exogeneity. *Econometrica* **51**, 277–304 (1983).
- MacKinlay, A. C. Event studies in economics and finance. *J. Econ. Lit.* **35**, 13–39 (1997).
- Al-Ubaydli, O. & List, J. A. in *Handbook of Experimental Economic Methodology* (eds Fréchette, G. R. & Schotter, A.) 420–462 (Oxford Univ. Press, New York, 2015).
- Harrison, G. W. & List, J. A. Field experiments. *J. Econ. Lit.* **42**, 1009–1055 (2004).
- Akerlof, G. A. Labor contracts as partial gift exchange. *Q. J. Econ.* **97**, 543–569 (1982).
- List, J. A. & Rasul, I. *Handbook of Labor Economics* Vol. 4 (eds Ashenfelter, O. & Card, D.) 103–228 (Elsevier, Amsterdam, 2011).
- Heckman, J. J. Causal parameters and policy analysis in economics: a twentieth century retrospective. *Q. J. Econ.* **115**, 45–97 (2000).
- Heckman, J. J. & Smith, J. A. Assessing the case for social experiments. *J. Econ. Perspect.* **9**, 85–110 (1995).
- Fréchette, G. R. & Schotter, A. *Handbook of Experimental Economic Methodology* (Oxford Univ. Press, New York, 2015).
- Cummings, R. G., Elliott, S., Harrison, G. W. & Murphy, J. Are hypothetical referenda incentive compatible? *J. Polit. Econ.* **105**, 609–621 (1997).
- Kramer, A. D., Guillory, J. E. & Hancock, J. T. Experimental evidence of massive-scale emotional contagion through social networks. *Proc. Natl Acad. Sci. USA* **111**, 8788–8790 (2014).
- Falk, A. & Heckman, J. J. Lab experiments are a major source of knowledge in the social sciences. *Science* **326**, 535–538 (2009).
- Maniadis, Z., Tufano, F. & List, J. A. One swallow doesn't make a summer: new evidence on anchoring effects. *Am. Econ. Rev.* **104**, 277–90 (2014).
- Levitt, S. D. & List, J. A. What do laboratory experiments measuring social preferences reveal about the real world? *J. Econ. Perspect.* **21**, 153–174 (2007).
- Pigou, A. C. *Theory of Unemployment* (Routledge, Abingdon, 2013).
- Marshall, A. *Principles of Economics* (Macmillan, London, 1890).
- Walras, L. *Éléments d'économie Politique Pure, Ou, Théorie de la Richesse Sociale* (F. Rouge, Lausanne, 1896).
- Al-Ubaydli, O., List, J. A. & Price, M. K. The nature of excess: using randomized treatments to investigate price dynamics. Preprint at <https://www.nber.org/papers/w16319> (2010).
- Al-Ubaydli, O. & List, J. A. *Handbook of Economic Field Experiments* Vol. 1 (eds Banerjee, A. V. & Duflo, E.) 271–307 (Elsevier, Amsterdam, 2017).
- Levitt, S. D. & List, J. A. Field experiments in economics: the past, the present, and the future. *Eur. Econ. Rev.* **53**, 1–18 (2009).
- Garraty, J. A. *Unemployment in History: Economic Thought and Public Policy* (Harper & Row, New York, 1978).
- Smith, V. L. An experimental study of competitive market behavior. *J. Polit. Econ.* **70**, 111–137 (1962).
- Nash, J. Non-cooperative games. *Ann. Math.* **54**, 286–295 (1951).
- Ramey, V. A. Can government purchases stimulate the economy? *J. Econ. Lit.* **49**, 673–685 (2011).
- Smith, V. L. Experimental auction markets and the Walrasian hypothesis. *J. Polit. Econ.* **73**, 387–393 (1965).
- Manning, A. An integration of trade union models in a sequential bargaining framework. *Econ. J.* **97**, 121–139 (1987).
- Shapiro, C. & Stiglitz, J. E. Equilibrium unemployment as a worker discipline device. *Am. Econ. Rev.* **74**, 433–444 (1984).
- Akerlof, G. A. The market for “lemons”: quality uncertainty and the market mechanism. *Q. J. Econ.* **84**, 235–251 (1970).
- Prendergast, C. The provision of incentives in firms. *J. Econ. Lit.* **37**, 7–63 (1999).
- Sparks, R. A model of involuntary unemployment and wage rigidity: worker incentives and the threat of dismissal. *J. Labor Econ.* **4**, 560–581 (1986).
- Akerlof, G. A. & Kranton, R. E. Identity, supervision, and work groups. *Am. Econ. Rev.* **98**, 212–17 (2008).
- Greenwald, B. C. Adverse selection in the labour market. *Rev. Econ. Stud.* **53**, 325–347 (1986).
- Spence, M. Job market signaling. *Q. J. Econ.* **87**, 281–306 (1973).
- Cain, G. G. *Handbook of Labor Economics* Vol. 1 (eds Ashenfelter, O. C. & Layard, R.) 693–785 (Elsevier, Amsterdam, 1986).
- Becker, G. S. *The Theory of Discrimination* (Univ. Chicago Press, Chicago, 1957).
- Nagin, D. S., Rebitzer, J. B., Sanders, S. & Taylor, L. J. Monitoring, motivation, and management: the determinants of opportunistic behavior in a field experiment. *Am. Econ. Rev.* **92**, 850–873 (2002).
- Boly, A. On the incentive effects of monitoring: evidence from the lab and the field. *Exp. Econ.* **14**, 241–253 (2011).

51. Holmstrom, B. & Milgrom, P. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *J. Law Econ. Organ.* **7**, 24–52 (1991).
52. Shearer, B. Piece rates, fixed wages and incentives: evidence from a field experiment. *Rev. Econ. Stud.* **71**, 513–534 (2004).
53. Bandiera, O., Barankay, I. & Rasul, I. Social preferences and the response to incentives: evidence from personnel data. *Q. J. Econ.* **120**, 917–962 (2005).
54. Shi, L. Incentive effect of piece-rate contracts: evidence from two small field experiments. *B. E. J. Econ. Anal. Policy* **10**, <https://doi.org/10.2202/1935-1682.2539> (2010).
55. Hong, F., Hossain, T., List, J. A. & Tanaka, M. Testing the theory of multitasking: evidence from a natural field experiment in Chinese factories. *Int. Econ. Rev.* **59**, 511–536 (2018).
56. Al-Ubaydli, O., Andersen, S., Gneezy, U. & List, J. A. Carrots that look like sticks: toward an understanding of multitasking incentive schemes. *South. Econ. J.* **81**, 538–561 (2015).
57. Benabou, R. & Tirole, J. Intrinsic and extrinsic motivation. *Rev. Econ. Stud.* **70**, 489–520 (2003).
58. Bertrand, M. & Duflo, E. *Handbook of Economic Field Experiments* Vol. 1 (eds Banerjee, A. V. & Duflo, E.) 309–393 (Elsevier, Amsterdam, 2017).
59. Bertrand, M. & Mullainathan, S. Are Emily and Greg more employable than Lakisha and Jamal? *Am. Econ. Rev.* **94**, 991–1013 (2004).
60. Gneezy, U., List, J. & Price, M. K. Toward an understanding of why people discriminate: evidence from a series of natural field experiments. Preprint at <https://www.nber.org/papers/w17855> (2012).
61. Neumark, D., Bank, R. J. & Van Nort, K. D. Sex discrimination in restaurant hiring: an audit study. *Q. J. Econ.* **111**, 915–941 (1996).
62. Pager, D. The mark of a criminal record. *Am. J. Sociol.* **108**, 937–975 (2003).
63. Kroft, K., Lange, F. & Notowidigdo, M. J. Duration dependence and labor market conditions: evidence from a field experiment. *Q. J. Econ.* **128**, 1123–1167 (2013).
64. List, J. A. The nature and extent of discrimination in the marketplace: evidence from the field. *Q. J. Econ.* **119**, 49–89 (2004).
65. Blaug, M. *The Methodology of Economics: Or, How Economists Explain* (Cambridge Univ. Press, Cambridge, 1992).
66. Akerlof, G. A. & Yellen, J. L. Fairness and unemployment. *Am. Econ. Rev.* **78**, 44–49 (1988).
67. Agell, J. & Lundborg, P. Theories of pay and unemployment: survey evidence from Swedish manufacturing firms. *Scand. J. Econ.* **97**, 295–307 (1995).
68. List, J. A. The behavioralist meets the market: measuring social preferences and reputation effects in actual transactions. *J. Polit. Econ.* **114**, 1–37 (2006).
69. Gneezy, U. & List, J. A. Putting behavioral economics to work: testing for gift exchange in labor markets using field experiments. *Econometrica* **74**, 1365–1384 (2006).
70. Pritchard, R. D., Dunnette, M. D. & Gorgenson, D. O. Effects of perceptions of equity and inequity on worker performance and satisfaction. *J. Appl. Psychol.* **56**, 75–94 (1972).
71. Loewenstein, G. & Schkade, D. in *Well-being: The Foundations of Hedonic Psychology* (eds Kahneman, D., Diener, E. & Schwarz, N.) 85–105 (Russell Sage Foundation, New York, 1999).
72. Lee, D. & Rupp, N. G. Retracting a gift: How does employee effort respond to wage reductions? *J. Labor Econ.* **25**, 725–761 (2007).
73. Kube, S., Maréchal, M. A. & Puppe, C. Do wage cuts damage work morale? Evidence from a natural field experiment. *J. Eur. Econ. Assoc.* **11**, 853–870 (2013).
74. Bellemare, C. & Shearer, B. Gift giving and worker productivity: evidence from a firm-level experiment. *Games Econ. Behav.* **67**, 233–244 (2009).
75. Levitt, S. D. & Neckermann, S. What field experiments have and have not taught us about managing workers. *Oxf. Rev. Econ. Policy* **30**, 639–657 (2014).
76. Fehr, E., Goette, L. & Zehnder, C. A behavioral account of the labor market: the role of fairness concerns. *Annu. Rev. Econ.* **1**, 355–384 (2009).
77. Fehr, E., Kirchsteiger, G. & Riedl, A. Does fairness prevent market clearing? An experimental investigation. *Q. J. Econ.* **108**, 437–459 (1993).
78. Al-Ubaydli, O. & List, J. A. Do natural field experiments afford researchers more or less control than laboratory experiments? *Am. Econ. Rev.* **105**, 462–66 (2015).
79. Card, D. & Krueger, A. B. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *Am. Econ. Rev.* **84**, 772–793 (1994).
80. Burda, M. C. A note on firing costs and severance benefits in equilibrium unemployment. *Scand. J. Econ.* **94**, 479–489 (1992).
81. Bazen, S. & Skourias, N. Is there a negative effect of minimum wages on youth employment in France? *Eur. Econ. Rev.* **41**, 723–732 (1997).
82. Endo, S. K. Neither panacea, placebo, nor poison: examining the rise of anti-unemployment discrimination laws. *Pace Law Rev.* **33**, 4 (2013).
83. Christiano, L. J., Eichenbaum, M. & Evans, C. L. Nominal rigidities and the dynamic effects of a shock to monetary policy. *J. Polit. Econ.* **113**, 1–45 (2005).
84. Duffy, J. *Experimental Macroeconomics* (Palgrave Macmillan, London, 2010).

### Competing interests

The authors declare no competing interests.

### Additional information

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Correspondence should be addressed to O.A.-U.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© Springer Nature Limited 2019