



# Hiring is Broken

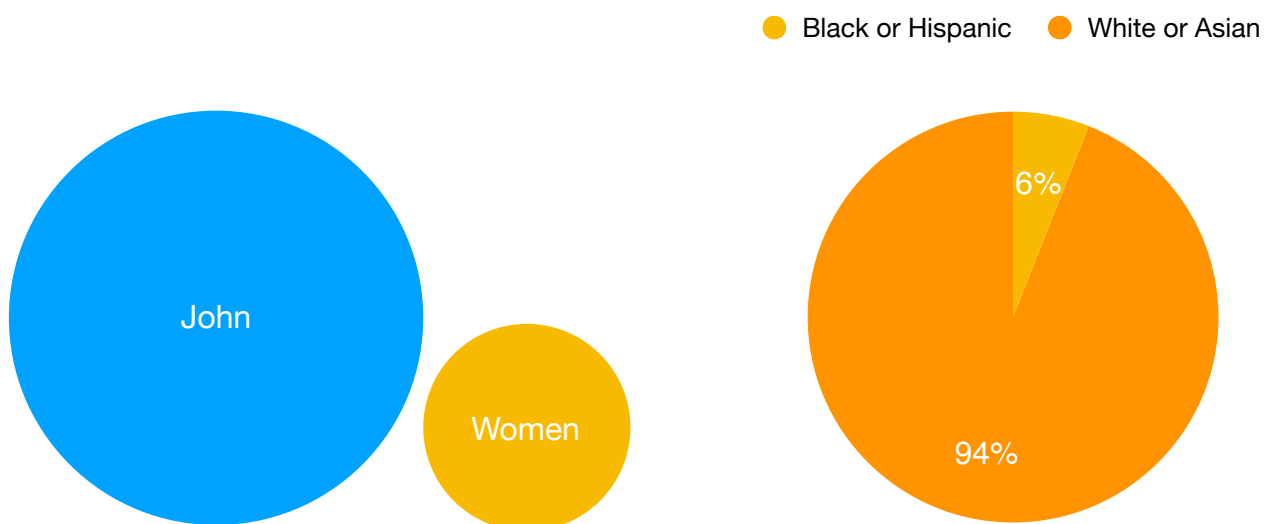


The Applied team

# Hiring is broken and the way we make decisions needs a radical overhaul

In August 2017 a 10-page memo by former Google software engineer James Damore ended up in public circulation. Damore's words stoked a furore after he claimed that women were biologically less well suited for jobs in tech and branded positive discrimination unfair.

Diversity remains a hot button issue at many of the world's top firms. Despite widespread diversity initiatives, there are still more FTSE 100 companies headed up by men called John than there are run by women. And just 6% of tech staff in Silicon Valley identify as Black or Hispanic.



But before we get into the thorny issues of positive discrimination and quotas, there are changes that hiring teams can make to promote diversity prior to, during and after the application process. And they have nothing to do with conventional pro-diversity initiatives.

Developed by behavioural scientists at Applied, this white paper outlines the six changes you can make to the way you hire that will improve outcomes for your business.

With a few small changes, recruiters can remove information that might cloud their views about a candidate, eliminate bias and make more objective candidate selections. They can also make sure they get better at attracting the right candidates in the first place and leave unsuccessful candidates with a more positive view of their company.

## About Applied

Applied is a platform that makes recruitment smarter, simpler and fairer. It's the brainchild of a team of behavioural and data scientists who are committed to using robust research to improve the hiring process. Applied is the first spin out of the UK's Behavioural Insights Team.

## Contents

1. Making more objective hiring decisions
2. Removing the risk of making impressionistic assessments
3. Crowdsourcing reviews for more accurate assessments
4. Writing more effective role descriptions and job adverts
5. Improving the way you communicate with unsuccessful candidates.

[Let's get started](#)



# Making more objective hiring decisions

Chances are you're not hiring the best people for the job. There, we said it. And as a result, your company is missing out on innovations, sales and improvements to products and services.

Organisations worldwide spend billions trying to attract and retain high performers because they are disproportionately productive. In knowledge-based industries, an employee in the top 1% is estimated to be 25 times as productive as the median employee.



So with all the money that's being spent on recruitment, why are we still not picking the right people every time?

Research shows the average recruiter takes fewer than six seconds to decide whether to call someone back for interview. Because six seconds is not long enough for the average brain to assimilate an entire CV, we scan for key details. We note where candidates went to school or university and form an idea about them. These quick judgements save us precious time. Unfortunately, researchers have found that they don't always result in optimal outcomes.

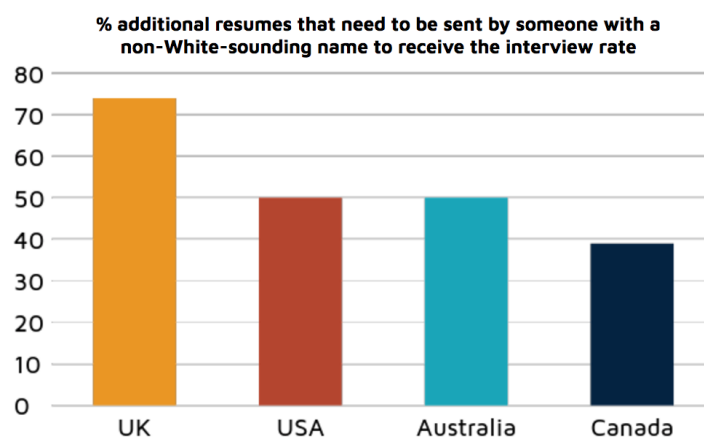
Whether we care to admit it or not, the way we interpret the world around us is shaped by implicit biases. These mental shortcuts often help us to make quick decisions in a complicated world, but in the context of recruitment, they can perpetuate race, gender, sexual identity, and socioeconomic inequality.

## The name game

Research shows that applicants with 'white-sounding' names do better, recruiters tend to favour candidates who they identify with and that it matters whose application gets read first. And that most recruiters make these decisions entirely unwittingly.

Experiments have shown that simply changing the name on an application can significantly alter call-back rates, even for employers with a public commitment to equal opportunity. And, despite the estimated USD8 billion spent annually on diversity training programmes, there's little to no evidence that they work.

Experiments in the US, Canada and Australia which involved sending fake resumes to real employers reveal that the response rates to otherwise identical applications can be up to 50% higher for 'White-sounding' names. In the UK, that figure is just over 75%.



## Perpetuating stereotypes

Women have a harder time applying to male-dominated industries and the reverse is also true. And studies also indicate that signalling homosexuality or lower socioeconomic class are detrimental to job prospects.

Established and systemic bias perpetuates industry stereotypes, which leads to an enduring lack of diversity across sectors. Many people's reaction to the assertion that there is systemic bias in the recruitment process is denial. And worryingly, experiments even show that men (and even more so male academics) are less convinced by the scientific evidence on gender discrimination. And that's part of the problem.

## Why diversity matters

The business case for diversity is now widely accepted. Diverse teams are associated with increased sales revenue, more customers, gains in market share, and more accurate pricing of assets. Diverse teams have also been shown to produce more influential research and be more innovative.

Why? The greater the diversity, the greater the number of perspectives. When teams are set up for diversity, that means more and different ideas are brought to the table, meaning teams don't approach problems in the same way, and this tends to make them better at solving complex problems. Put simply, teams that are set up for diversity can benefit from the avoidance of 'groupthink'. Yet despite decades of diversity initiatives spanning

everything from affirmative action to quotas, we are far from solving the problem.

Unfortunately for the well-meaning employer, simply thinking that you're objective, committing to diversity publicly, or even investing in diversity training programmes is no guarantee you will achieve it. It's estimated that USD8 billion is spent by US corporates each year on diversity training programmes, and yet there's little to no evidence that they actually change outcomes. Worse still, some studies have shown that they can backfire. And studies have shown that companies that are openly committed to diversity are as likely to discriminate as those who aren't.

It's not that people don't care, or don't want to do better. But years of research tell us that rewiring the brain is painstaking work, and a thousand times harder to do when your environment doesn't change with you. In the same way that committing to a weight loss regime is hard when your house is filled with chocolate, being less biased is near impossible when the systems your HR team rely on don't support you.



## Hacking the system

A few years ago we decided to change the way we processed job applications. We started manually blinding irrelevant information from the application. After eight hours of painstaking work with a permanent marker, we started to appreciate why no one else was doing it: it was time-consuming, error-prone and frankly, a bit of a hassle.

So we started working on a platform that did the heavy lifting for us and Applied was born. Applied automatically removes irrelevant information meaning reviewers can concentrate on the information that really matters.

Not only does stripping out irrelevant information like a candidate's name eliminate bias, but it also helps save time.

## Removing the risk of making impressionistic assessments

We quickly realised that while anonymising applications is important, it's just the start.

After extracting irrelevant information that can distort initial applicant selection, we got closer to making more objective hiring decisions. But what about a recruiter's ability to score candidate responses consistently?

Our research shows when reading applications from start to finish, information provided in one answer can influence the way a reviewer perceives subsequent answers. For example, a great answer to question one

can create a 'halo effect', which inflates the perceived quality of the next answer.

Not only that, but the order of candidates matters. In much the same way as research has shown that who you vote for, who you find attractive, or even how you perceive colours can be affected by the order in which options are presented, it turns out that how we score people is also context-dependent.

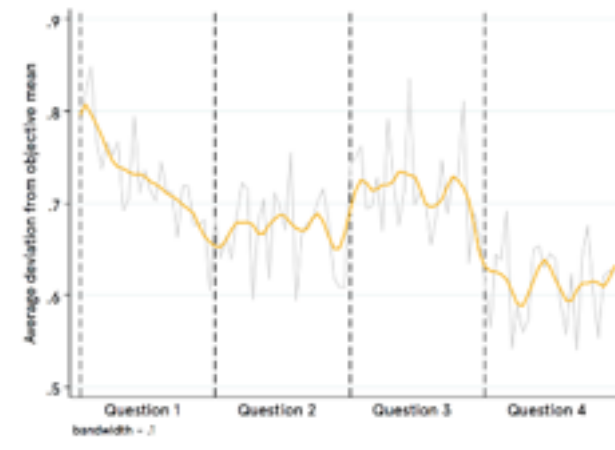
In 2016, we invited around 150 reviewers on an online research platform to rate the responses by 80 candidates to four work-related challenges. Each reviewed 100 unique responses, drawn from real candidates who had previously applied to a position.

Candidates' responses were anonymised, chunked by question, and the order of their responses within each question was randomised for each reviewer. We then compared these to a benchmark score combining everyone's opinion of that response.

## **The results:**

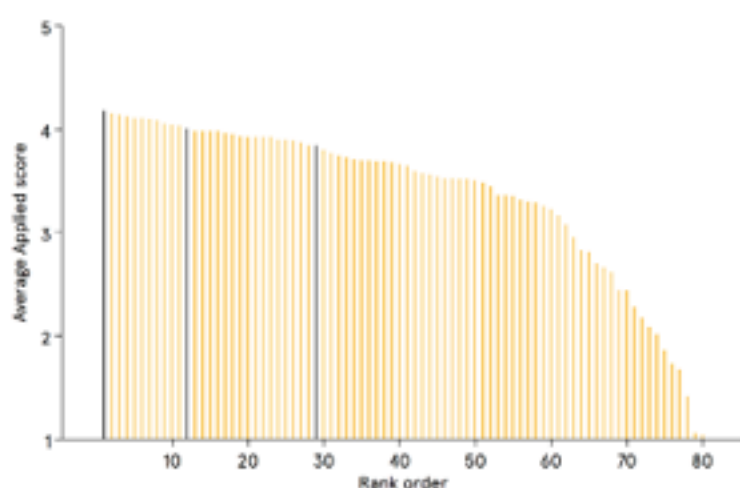
### **Reviews become more accurate over time. (Finding 1)**

Our research showed that the order in which the answers were read affected how they were scored. Since we'd randomised across lots of reviewers, we could compare whether the score given to a particular response was different if the reviewer read it as the first in the batch, or 9th, 17th, or last.



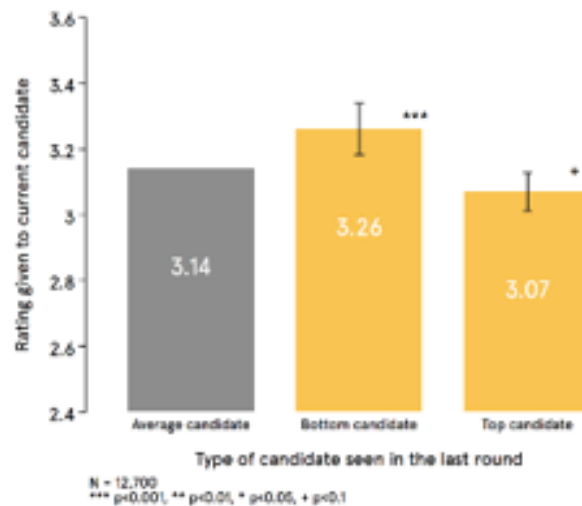
## There's an advantage to being first. (Finding 2)

We also found that being first was an advantage: The average rating across all candidates was 3.35, but being rated first increased that to 3.52. While those effects don't sound big, that's the same gap as being ranked 1st or 12th (borderline for interview), or 12th and 29th (rejected for interview).



## It matters who comes before you. (Finding 3)

An average candidate appears worse (and gets a lower score) if they come after a phenomenal candidate, but they'll seem far better (and get a higher score) if they come after a poor one.



We took the top and bottom ten candidates on each question (as rated by all reviewers) and looked at what impact seeing that response had on the scores for the next few candidates.

To make sure reviewers focus on the right things, we added two additional features to Applied.

First, we chunk up applications by question so reviewers judge each response in turn, instead of each candidate in turn. Blinded horizontal comparison makes it easy for reviewers to identify the best responses and shields against impressionistic judgements and 'halo effects'.

Second, our review algorithm makes sure that candidate responses are randomised across each question and all reviewers. So no response is read in the same order. This quite simple design choice turns out to make a world of difference to making sure that no-one is advantaged or disadvantaged by the order in which they're assessed.

## Crowdsourcing reviews for more accurate assessments

We know that individuals working on their own are fallible. Part of the reason that organisations lack diversity is we tend to hire people in our own image (affinity bias). And everyone has their own version of what 'good' looks like, which can result in wildly different shortlists.

By contrast, decades worth of research illustrate that soliciting multiple opinions can reduce the risk of error. Studies have shown that crowds will beat experts when they are:

- Diverse – meaning they bring varying degrees of knowledge and insight;
- Independent – that is, individuals' opinions aren't affected by those around them;
- Decentralised – meaning they are able to specialise and draw on local knowledge; and
- Aggregated – which is to say there's a mechanism for collecting views and converting them into collective intelligence.

When these conditions hold, there are very few experts who will outperform the group. In fact, researchers have even shown that US defence intelligence analysts with access to classified information can be beaten by some rudimentarily-educated amateurs: largely because they come to conclusions too quickly and struggle to update



their opinions in the face of new and conflicting information.

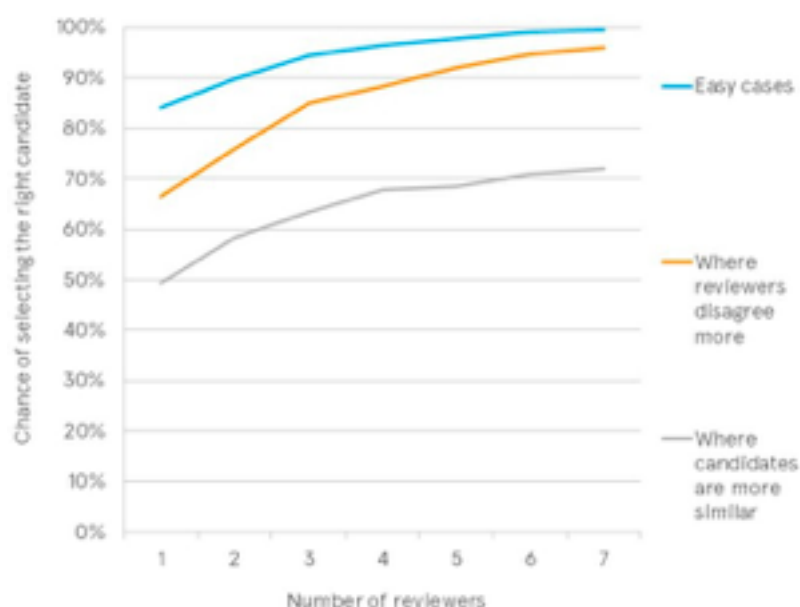
So we ran experiments to test how many people you need to review candidates to make sure you get a robust measure of their quality.

Through a simple online experiment we asked about 400 reviewers to rate the responses of four hypothetical (unnamed) candidates to a generic recruiting question.

Unsurprisingly, their combined ratings easily identified the best response. But most organisations can't afford to ask hundreds of people to help them select a candidate.

So we sought to discover the optimal number of reviewers to involve in the application process.

We took our data and ran statistical simulations to estimate the probability that different groups could correctly select the best candidate. We created 1,000 combinations of reviewers in teams of different sizes, ranging from one to seven people. We then pooled them by the size of the group and averaged their chance of selecting the right candidate.



The following graph confirms that with more people, you are more likely to correctly identify the best person.

## **But what happens when people disagree over what 'good' looks like?**

That's what the orange line shows. Where reviewers disagree more on a given candidate you need to pool more opinions to gain the same level of judgmental accuracy. Moving from one to three reviewers has a big impact: you go from a one in three chance of getting it wrong to 15%.

Finally, what about situations where candidates are really similar, and it's hard to distinguish between them? The grey line reveals what we found when we tested the crowd's ability to separate the second and third best candidates (whose responses were similarly good). We found crowds are even more important here, and one person working on their own can be trusted no more than a coin toss!

## **The importance of independence**

Our study showed that combining the views of different reviewers results in a more accurate shortlisting process. But it's worth emphasising that it's not just a matter of asking a group of three people to make the decision together.

Humans are social animals, and how we combine our diverse opinions is critical. If not managed appropriately, we can fall prey to collective decision making errors like

'groupthink', social hierarchy bias, and social conformity bias. We're all affected by the views of those around us, and that can mean we question our true opinions, or we say what we think others want to hear. For this reason, what we showed in our experiment was the outcome of independent opinions of reviewers. We can expect that there would have been even worse errors committed if we'd orchestrated reviewers to scoring the candidates as a group.

Our research revealed that three independent reviewers generates the fairest and most accurate assessments. Reviewers can be easily selected, randomised and comparative assessments can be assigned. And the benefit is that getting independent assessments results in both better judgements and it's more efficient, avoiding the need to coordinate diaries and find time to reach consensus. The platform can then conduct detailed analytics on how people score, what skill sets they're great at identifying, and what questions result in highly subjective or objective reviews.

## Writing more effective role descriptions and job adverts

Job adverts matter. Although an employer's brand and reputation will be one of the most important predictors of whether someone applies for a job, it's not the whole story. The way you frame the advert and the words you use will affect the types of candidates who apply.

One study found that specific words are more likely to attract female or male candidates. Research shows job adverts containing stereotypically masculine-coded words are less appealing to women than stereotypically feminine-coded descriptions. And other work points to correlations between gender, confidence, and how willing candidates are to 'take a punt' even when they don't have all the required characteristics.

Applied comes with two built-in text analytics tools. First, we've embedded a gendered language tool into our platform that automatically detects masculine or feminine coded words or language within a job description and lists these to avoid inadvertent bias. Since we also manage selection and hiring and have detailed diversity data, we also dynamically update the tool over time to capture the latest data in how words affect attraction.

Gender Balance:  
Strongly Feminine-coded

Consider alternatives to these words

decisions	understand
competing	collaboration
leads	responsibilities
leads	support
leads	responsibilities
leads	tenders)
	supporting
	support
	support
	supporting
	interpersonal
	understanding

A second tool gives job descriptions a readability score, which enables hiring teams to make sure their adverts are inclusive and easy to understand. Just as recruiters are notorious for making quick decisions, so are prospective candidates. Some industry research indicates candidates scan an advert for as little as 10 seconds, so our readability score is designed to also help you write in a way that'll improve engagement.

Reading difficulty: 15 years

You've scored 62.5 / 70 (where 70 is the optimal score) on the Flesch reading ease test.

This means your expectation of applicants' reading ease for this job description is similar to that of Reader's Digest .



## Improving the way you communicate with unsuccessful candidates

Industry research indicates that as many as one in four candidates walk away from a recruitment process with an actively bad impression of the organisation. And of those, one in three will tell their friends, and 12% will post on social media.

For consumer-facing organisations, the impact could be more acute. The same research shows that almost one in five people who've had a bad candidate experience will stop using products/services from that organisation in the future. From which they estimate this could amount to a foregone revenue cost in the millions each year.

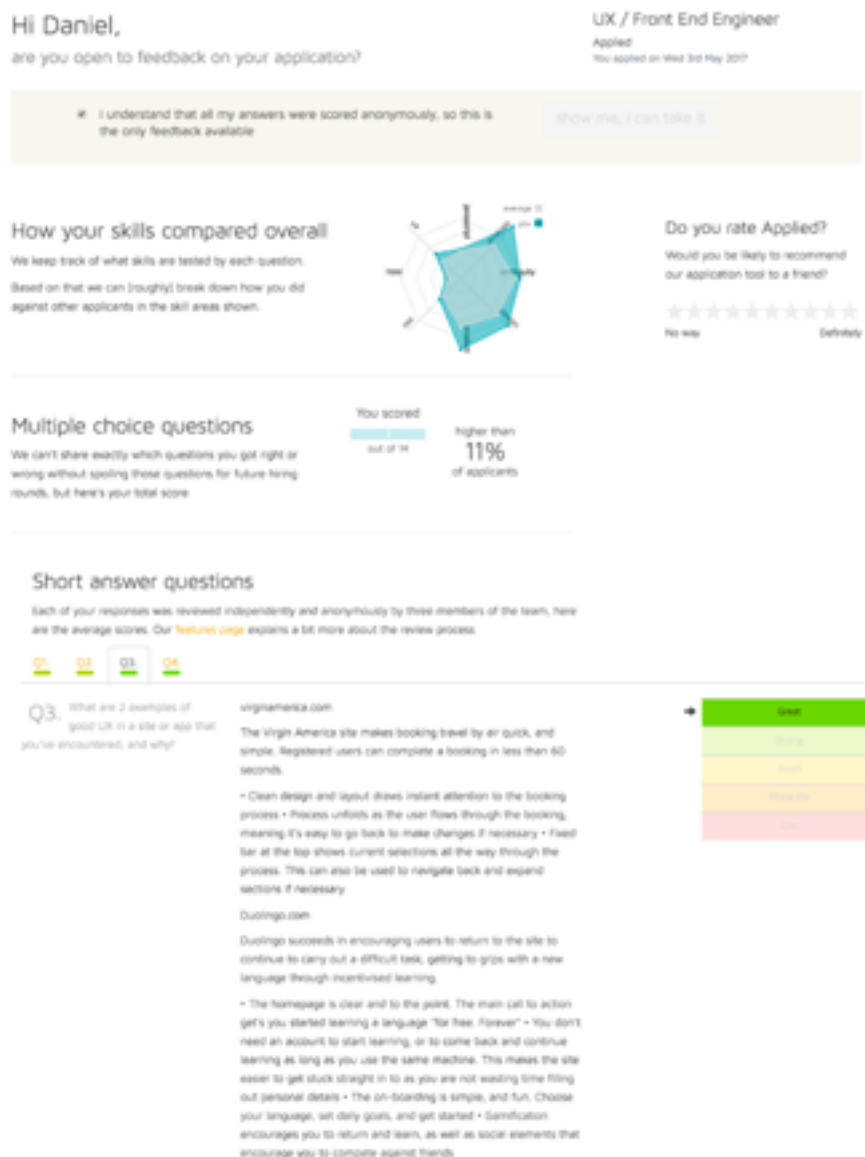
There's a business case for providing candidate feedback as 'employer brands' become more relevant in today's job market. But hiring managers find it almost impossible to provide personalised, insightful feedback to each candidate.

For most people it comes down to the time cost - collating evidence as to why someone wasn't right for the job can be time consuming. Most recruiters simply don't have the time to follow up to offer one-on-one feedback to hundreds of disappointed candidates.

We wanted to try to bridge this somehow, and so we decided to take the (apparently radical) step of making it the default that when someone is sent an email telling them they haven't progressed, that they get a link to a

webpage they can go to (if they want to) to find out how they were scored (and the process behind it).

Here is the feedback page that went out to a recent hire into Applied.



Applied doesn't reveal all the details about how questions are scored, but it does aim to give candidates a sense of how they compared to others who applied and where their strengths and weaknesses are.

Generally the feedback has been positive. While giving candidates feedback, we also asked them to rate the process. Of the almost 1,000 candidates who did so, one in three gave a rating of 10/10, and the median score they gave was 9/10. Clearly, candidates value meaningful feedback.

## What all this means

Applied was created by a team of behavioural and data scientists driven by the goal of improving the fairness, quality and accuracy of the hiring decisions we make. From research conducted by external teams and ourselves we've been able to identify a few of the ways in which we know unconscious biases can impact on our decisions.

By moving beyond traditional approaches and using the latest research insights, companies have the ability to simplify and streamline recruitment. And build better performing teams across their companies as a result.

Follow our blog for the latest thinking on how insights from behavioural science can help your company.

Blog: [medium.com/finding-needles-in-haystacks](https://medium.com/finding-needles-in-haystacks)

Email: [hello@beapplied.com](mailto:hello@beapplied.com)

[beapplied.com](https://beapplied.com)

# References

1. O'Boyle, E. & Aquinis, H., (2012), 'The best and the rest: Revisiting the norm of normality of individual performance', *Personnel Psychology*, 65, pp. 79-119.
2. Kahneman, D., (2011), *Thinking, Fast and Slow*, Penguin.
3. Bertrand, M. & Duflo, E., (2016), 'Field Experiments on Discrimination', Prepared for the *Handbooks of Field Experiments*; Bohnet, I., (2016), *What Works: Gender Equality by Design*, Harvard University Press; Neumark, D., (2016), 'Experimental Research on Labor Market Discrimination', NBER Working Paper 22022; Behavioural Insights Team, (2015), 'A Head for Hiring: The Behavioural Science of Recruitment and Selection', [behaviouralinsights.co.uk/publications/a-head-for-hiring-the-behavioural-science-of-recruitment-and-selection/](http://behaviouralinsights.co.uk/publications/a-head-for-hiring-the-behavioural-science-of-recruitment-and-selection/).
4. Kang, S. et al, (2016), *Whitened Résumés: Race and Self-Presentation in the Labor Market*, *Administrative Science Quarterly*, 61, 3, 469-502.
5. Kalinoski, Z. et al, (2013), A meta-analytic evaluation of diversity training outcomes, *Journal of Organizational Behavior*, 34, 8, 1076-1104; Dobbin, F. & Kalev, A., (2016), 'Why Diversity Programs Fail And what works better', *Harvard Business Review*, 94, 7-8, 52-60.; Kalev, A., Dobbin, F., & Kelly, E., (2006), Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies, *American sociological review*, 71, 4, 589-617; Bohnet, I., (2016), *What Works: Gender Equality by Design*, Harvard University Press.
6. Bertrand, M. & Mullainathan, S., (2004) 'Are Emily and Greg More Employable Than Lakisha and Jamal - A Field Experiment on Labor Market Discrimination', *American Economic Review*, 94, 4, 991-1013; Booth, A. & Leigh, A (2010), 'Do Employers Discriminate by Gender? A Field Experiment in Female-Dominated Occupations', *Economics Letters*, 107, 236-238; Oreopoulos, P., (2011), 'Why Do Skilled Immigrants Struggle in the Labor Market?, A Field Experiment with Thirteen Thousand Resumes', *American Economic Journal: Economic Policy* 3 (November 2011): 148-171.
7. Wood, M. et al, (2009), 'A test for racial discrimination in recruitment practice in British cities', report for the UK Department of Work and Pensions, Research report No. 607.
8. Bohnet, I., (2016), *What Works: Gender Equality by Design*, Harvard University Press.
9. Rule, N. et al, (2016), 'Subtle Perceptions of Male Sexual Orientation Influence Occupational Opportunities', *Journal of Applied Psychology*, 101, 12, 1687-1704; Rivera, L. & Tilcsik, A. (2016), 'Class Advantage, Commitment Penalty, The Gendered Effect of Social Class Signals in an Elite Labor Market', *American Sociological Review*, 18, 6;
10. Handley, I. et al, (2015), 'Quality of evidence revealing subtle gender biases in science is in the eye of the beholder', *Proceedings of the National Academy of Sciences of the United States of America*, 112, 43, 13201-13206



11. Gompers, P. & Wang, S., (2017), 'And the Children Shall Lead: Gender Diversity and Performance in Venture Capital', NBER Working paper 23454; Bohnet, I., (2016), What Works: Gender Equality by Design, Harvard University Press; McKinsey & Co, (2014), Diversity Matters. McKinsey & Co, (2014), Diversity Matters; Phillips, K., (2014), 'How diversity makes us smarter', Scientific American, <http://www.scientificamerican.com/article/how-diversity-makes-us-smarter/>; Dow Jones (2012), 'Women At The Wheel: Do Female Executives Drive Start-Up Success?'; <http://www.scientificamerican.com/article/how-diversity-makes-us-smarter/>.
12. Galinsky, A. et al, (2015), 'Maximizing the Gains and Minimizing the Pains of Diversity A Policy Perspective', Perspectives on Psychological Science, Vol 10, No. 6, pp. 742-748.
13. Uhlmann, E. & Cohen, C., (2007), "'I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination', Organizational Behavior and Human Decision Processes, 104, 207-223; Kaiser, C. et al, (2012), 'Presumed Fair: Ironic Effects of Organizational Diversity Structures', Journal of Personality and Social Psychology, 104, 3, 504-519.
14. Bohnet, I., (2016), What Works: Gender Equality by Design, Harvard University Press.
15. Kang, S. et al, (2016), Whitened Résumés: Race and Self-Presentation in the Labor Market, Administrative Science Quarterly, 61, 3, 469-502.
16. Wedell, D., Parducci, A. & Geiselman, R. (1987), 'A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation', Journal of Experimental Social Psychology, 23, 3, 230-249; Miller, J. & Krosnick, J., (1999), 'The Impact of Candidate Name Order on Election Outcomes', Public Opinion Quarterly, 63, 2, 291-330; Thaler, R., Sunstein, C. & Balz, J., (2010), 'Choice architecture', [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1583509](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1583509); NASA Ames Research Centre, [https://colorusage.arc.nasa.gov/Simult\\_and\\_succ\\_cont.php](https://colorusage.arc.nasa.gov/Simult_and_succ_cont.php)
17. <https://medium.com/finding-needles-in-haystacks/hiring-honeybees-and-human-decision-making-33f3a9d76763>
18. Tetlock, P. & Gardner, D., (2015), Superforecasting: The Art and Science of Predictions, Crown.
19. See a full discussion in Bohnet, I., (2016), What Works: Gender Equality by Design, Harvard University Press.
20. CEB, <https://www.cebglobal.com/human-resources/recruiting/candidate-experience.html>.