Recruiters prefer expert recommendations over digital hiring algorithm: a choice-based conjoint study in a pre-employment screening scenario

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Management Research Review</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>MRR-06-2020-0356.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Original Article</td>
</tr>
<tr>
<td>Field Categories:</td>
<td>Human Resource Management</td>
</tr>
<tr>
<td>Keywords:</td>
<td>hiring, algorithm, gender, LinkedIn, choice-based conjoint, technology</td>
</tr>
</tbody>
</table>
Title: Recruiters prefer expert recommendations over digital hiring algorithm: a choice-based conjoint study in a pre-employment screening scenario

Abstract
Purpose: The present study analyzes aspects of decision making in recruitment. Using a choice-based conjoint experiment with typified screening scenarios, it was analyzed what aspects will be more important for recruiters: the recommendation provided by a hiring algorithm or the recommendation of a human co-worker; gender of the candidate and of the recruiter was taken into account. Methodology: A total of 135 recruitment professionals (67 female) completed a measure of sex roles as well as a set of 20 choice-based conjoint (CBC) trials on the hiring of a pharmacologist. Findings: Participants were willing to accept a lower algorithm score if the level of the human recommendation was maximum, indicating a preference for the co-worker’s recommendation over that of the hiring algorithm. Neither the biological sex of the candidate nor of the participant influenced in the decision. Research limitations: Participants were presented with a fictitious scenario that did not involve real choices with real consequences. In a real-life setting, considerably more variables influence hiring decisions. Practical implications: Results show that there are limits on the acceptance of technology based on artificial intelligence in the field of recruitment, which has relevance more broadly for the psychological correlates of the acceptance of technology. Originality/value: Results show that there are limits on the acceptance of technology based on artificial intelligence in the field of recruitment, which has relevance more broadly for the psychological correlates of the acceptance of technology. An additional value is the use of a methodological approach (CBC) with
high ecological validity that may be useful in other psychological studies of decision-making in management.

**Keywords**: choice-based conjoint; recruitment; hiring algorithm; artificial intelligence; gender role; LinkedIn

**Article classification**: Research paper
Introduction

The needle in the haystack: searching online for the best candidate

The increasing digitalization of society is changing how human resources departments hire, monitor and also dismiss employees. As Heric (2018) emphatically writes:

“In the ongoing battle for top talent, digital technologies play an increasingly prominent role. Seismic changes in the workforce and workplace, driven by demographics, alternative work arrangements and various distance-busting technologies, are rewriting the rules of engagement. While digital tools accelerate that shift, they also offer new ways for human resources teams to attract great talent, then motivate and equip them.” (Heric, 2018, p.1).

This process is not restricted to human relations departments seeking for talent or highly skilled employees. Digital tools have gradually Online information has already become an important integral part of most the hiring processes during the last two decades, both for job candidates and recruiters (Roulin and Bangerter, 2013; El Ouiridi et al., 2016). During the initial process of hiring personnel, much of the information about potential candidates is now frequently extracted from online sources. For example, on LinkedIn, recruiters can view not only a traditional CV, but also a candidate’s contacts, recommendations and personal information (Zide et al., 2014). However, online sources offer not only explicit information provided by the profile owner (studies, languages, etc.) but also implicit information. According to some studies (Back et al., 2010; Azucar et al., 2018) these psychological and psychosocial characteristics can be inferred to a high degree of accuracy in social networks such as Facebook (Kluemper et al., 2012), but also in professional networks such as LinkedIn (van de Ven et al., 2017), in which users present themselves in more standardized ways.
Companies can use this implicit information to judge the fit between a candidate and a position, not only in terms of training, experience, skills and personality but also in terms of the company’s culture, values and climate. However, this implicit information presents significant problems. First, and despite the “remarkable accuracy” (Azucar et al., 2018) of raters’ prediction of psychosocial characteristics based on digital traces, the recruiters’ impressions may be strongly biased by their personal preferences and not be a good predictor of the candidate’s future performance. Bias has always exerted a degree of influence on the hiring process, but online networks have the risk of introducing bias into the hiring process at a very early stage and even before the first interview (Caers and Castelyns, 2011). Second, the large volume of accessible information has also complicated hiring: finding the best candidate among many is increasingly like finding the famous needle in the haystack. To combat the problems of human bias and data overload, an innovative mechanism called the hiring algorithm has been championed as a possible solution.

**Algorithmic versus intuitive methods in recruitment**

Recruiters are expected to take into account exclusively the candidates’ proficiency related to their future professional performance (for example, training, experience, skills for the specific job, etc.) and therefore are urged to rely on analytics rather than on intuition (Davenport, 2006). When recruiters apply their intuition, they unwittingly consider factors that may be irrelevant to later job performance. Factors that affect hiring decision include candidates’ age and sexual orientation (Drydakis, 2009), weight (Swami et al., 2008), facial attractiveness (Tews et al., 2009), race and sex (Branscombe and Smith, 1990), and, last but not least, the recruiter’s own sex and national culture (El Ouirdi et al., 2016). Recruiters give subjectively desirable
applicants more favorable evaluations than subjectively undesirable applicants (Chiang and Suen, 2015).

In order to reduce these biases in the prediction of employee performance, decision aids have been developed over the decades. In an early meta-analysis (Vinchur et al., 1998), paper-and-pencil tests alone were already shown to outperform the unstructured interview. In more advanced analytical or algorithmic procedures, each piece of information is given a numerical value; these values are then combined to arrive at an overall score, and therefore a well-informed decision can be made on the candidates’ suitability. Finally, holistic methods (expert, intuitive, subjective judgment) include individual judgments and group consensus (Kuncel et al., 2013). However, traditional job interviews have been found to have almost no incremental predictive effect above and beyond standardized tests of ability and personality (Cortina et al., 2000). In a meta-analysis of 17 studies comparing these two methods (Kuncel et al., 2013), the authors conclude that, although the holistic method is preferred in business and academic contexts, this method overlooks relevant information for predicting job performance. Despite the evidence, there is still significant resistance to standardized hiring procedures (Lievens and De Paepe, 2004; Highhouse, 2008; Lodato et al., 2011).

Artificial intelligence (AI) and hiring algorithms add a new element into the hiring process. AI refers to a broad spectrum of technologies that allow a computer to perform tasks that normally require human cognition, including decision-making (Capelli et al., 2018). Machine learning (an aspect of AI) enables technology to create advanced new based on previous experience. Some companies offering predictive hiring algorithms also claim that these algorithms can determine the degree of compatibility between a candidate and a future supervisor. Start-ups have been using recent advances in machine learning to develop this kind of technology, and more and more companies
are relying on digital algorithmic decision-making. They do this based on the assumption that hiring recommendations made by this highly advanced technology is both efficient and impartial - at least more so than any human recruiter. Beyond providing standardized information about education and job experience, AI can scan social media data to assess the candidate’s values, beliefs, and attitudes (Upadhyay and Khandelwal, 2018). A human, in contrast, evaluates information selectively and may even suppress relevant information when faced with other information that subjectively has very high value (for example, the candidate’s sex). In theory, subjective bias should be eradicated with appropriately programmed algorithms. Some authors (Kuncel et al., 2013) come to a clear and simple conclusion that "algorithm beats instinct". According to a survey (see Lewis, 2018), 63 percent out of 770 talent acquisition professionals affirmed that AI had changed the way of recruiting, and that with AI higher-quality candidates can be found. However, a totally bias-free hiring process based on AI is difficult to achieve, as seen with Amazon in 2015, where the recruiting system configured machine-learning algorithms that ended up downgrading certain (female) resumes (see Lewis, 2018). The fast advances in AI are already dealing with this kind of shortcomings, and according to Pricewaterhouse-Coopers (2017), 40 percent of the HR functions of international companies (mainly USA and Asia) are currently using AI-applications. Still, many recruiters seem to be reluctant to embrace this technology, as, in their opinion, an algorithm can never replace human empathy and intuition.

Technology and sex/gender

This leads us to the question of which people rather “trust the computer”, based on pre-installed and objective criteria, and which people prefer to rely on the hiring recommendation made by other humans, despite the unreliability of intuition. Previous
research suggests that one of these factors is the decision maker’s sex. Using the technology acceptance model (TAM) as a framework, we analyzed psychological gender-related variables associated with reliance on hiring algorithms. TAM, in its original (Venkatesh and Davis, 2000) and expanded version (Venkatesh et al., 2000; Venkatesh and Bala, 2008), has been widely used to explain the potential user’s behavioural intention to use a technological innovation and proven to be a valid and robust model (King and He, 2006). Gender differences in technology acceptance and behavioural intentions have been found to show that, for instance, men have a higher level of intention to adopt mobile health systems than women (Zhang et al., 2014).

Men and women also adopt different decision-making processes in evaluating new technologies (Venkatesh et al., 2000). This difference stems from the finding that technology use is more congruent with the gender role of men (Elsbach and Stigliani, 2018); men have been found to have a more favourable attitude toward technology (Cai et al., 2017) and women are more likely to experience technophobia (Gilbert et al., 2003). In summary, technology acceptance and use seem to be more challenging for women (Teo et al., 2015; Sobieraj and Krämer, 2019). These differences might also apply to the more specific area of AI. Therefore, this paper will investigate whether female recruiters, compared to male recruiters, valued the source of information (algorithm vs human) about candidates differently.

However, such sex differences seem to be domain-specific. For instance, no gender effect was found in regards to social networking tools and online video sharing tools (Huang et al., 2013). Also, no gender effect was found in the context of healthcare for the elderly (Wong et al., 2012). Taking into account these contradictory findings, we agree with Venkatesh, Morris & Ackermann (Venkatesh et al., 2000) who suggest that it might be rather gender role (based on the construct of masculinity and femininity)
than biological sex that influences technology acceptance. However, with both male and females now being born within a technological era, these social constructs should slowly dissipate.

The gender role construct has a long history. Social role theory (Eagly, 1987) provides an explanation for gendered behavior. Sandra Bem (Bem, 1981) emphasized the role of culture, proposing the Bem Sex Role Inventory to assess the extent to which the culture’s definitions of desirable female and male attributes are reflected in an individual’s self-description. Both the descriptive and prescriptive aspects of gender roles are well documented (Eagly and Karau, 200220): Gender roles are not simply descriptive or explanatory categories, but also prescriptive, referring to people’s perceptions as to what they are expected by others to be and how they are expected to behave (Oberst et al., 2016). Importantly, Eagly & Karau argue that the majority of these beliefs about men vs women pertain in fact to communal vs. agentic attributes.

Men are expected to be more agentic and women to be more interpersonally oriented. However, the self-attribution of gender stereotypes has changed over the past decades. Whereas the feminine role has traditionally had negative connotations, femininity is now considered more desirable, depending on the context and country (Gartzia and Lopez-Zafra, 2014).

While in early conceptualizations, masculinity-femininity was considered a continuum, i.e. a bipolar dimension, later theories and research shifted from the bipolar alternatives of being masculine or feminine to a quadripolar typology in which sex roles could develop as masculine and feminine, masculine and not feminine, feminine and not masculine, or neither masculine nor feminine (Heilbrun, 1976). Since then, most instruments assessing gender roles use independent measures of two dimensions (masculinity and femininity). Recently, the expressions “masculinity” and “femininity”
have been replaced by less gender-stereotypical expressions. As a result, “communal”, also called “expressive” (Hentschel et al., 2019) characteristics, which are attributed more strongly to women, are described primarily as a concern for the welfare of other people (for example, being affectionate, helpful, kind, sympathetic, interpersonally sensitive, nurturing, and gentle).

In contrast, “agentic” - or “instrumental” - (Hentschel et al., 2019) characteristics, which are attributed more strongly to men, are described primarily tending to be assertive, controlling, and confident. People with more affiliative needs and communal characteristics may be more reluctant to use and rely on technology when it comes to making decisions about people.

Purpose of the present study

The present study aims at investigating how the source of information about candidates (a hiring algorithm or human expert) and participants’ sex and gender role affects their hiring decisions. To do so, we conducted an experiment using choice-based conjoint methodology (CBC) with typified fictitious scenarios to discover whether, in the early phases of the pre-employment screening process, recruiters were most influenced by the recommendation provided by an algorithm, by the recommendation provided by a person that could include objective and subjective aspects or the sex of the candidate. We tracked whether this preference was associated with the biological sex of the recruiters and / or their gender role (in terms of instrumentality and expressiveness). We expected to find that, in a fictitious hiring scenario, recruiters would give higher credit to recommendations made by human experts than to recommendations based on artificial intelligence (hypothesis 1). Furthermore, women would give recommendations made by human experts more often than men (hypothesis 2). Finally, we expected to
find a relationship between instrumentality and the preference for a recommendation made by the algorithm, and between expressiveness and the preference for personal recommendations (hypothesis 3).

Methodological considerations

Because the use of self-report questionnaires in psychological research presents some problems, mainly due to social desirability, bias can influence both paper-and-pencil and computer-based questionnaires (Dodou and De Winter, 2014). This, is even more, the case in situations of decision making, where we do not know how the decision-maker would decide in real situations. Therefore, many studies in organizational decision-making use experimental and simulation techniques (Koch et al., 2015; El Ouirdi et al., 2016), in order to rule out the possible influence of social desirability in the responses. In this study, in order to achieve more reliable responses and to detect possible biases of the participants, a relatively novel experimental procedure was used: the simulation of scenarios in decision-making using choice-based conjoint (CBC). CBC is a method developed in market research that allows researchers to analyze how people make decisions about something that is offered to them (a service, product or scenario). It requires research participants to make a series of trade-offs; the analysis of these trade-offs reveals the relative importance of the different attributes of the offer. This evaluation is usually done by asking a sample of the population to indicate their preferences regarding a succession of possible offers, defined by specific attributes, each with different levels. The subsequent statistical analysis makes it possible to determine which combination of these attributes maximizes the probability of choice. CBC is similar to real experiments insofar as it randomly assigns tasks or treatments to participants who reveal their preferences for
attributes, without explicitly stating or ranking them as in a traditional questionnaire. Participants do not even need to be aware of their preferences (Giersch and Dong, 2018).

In addition to its use in market studies, CBC has also been incorporated into a wide variety of different research areas. For example, in education research to study decision-making in school contexts (Blain-Arcaro et al., 2012), in discrimination studies to detect covert attitudes that contradict overt choices (Caruso et al., 2009); in psychological studies about choices on social media (Carbonell and Brand, 2018); in patient-centred decision-making in healthcare (Wilson et al., 2014; Flöthmann et al., 2018; Fraenkel et al., 2018); and finally, in selection and hiring processes (Connole et al., 2014; Flöthmann et al., 2018).

Method

Participants

We contacted 449 male and female recruiters (either human resources professionals working inside a company or headhunters working externally) through LinkedIn Premium, utilizing a personalized message and invited them to participate in the study. 193 agreed to participate.

Instruments

Pre-selection scenarios

Using specialized software for conjoint analysis (Lighthouse Studio 9.6.1 Sawtooth [www.sawtoothsoftware.com]), we devised 20 pre-selection scenarios about hiring a pharmacologist for a multinational pharmaceutical company. This job description was intended to be gender-neutral; although pharmacy is a STEM career, in
Spain it has a high number of female students, and over 50% of jobs in the pharmaceutical industry in Spain are occupied by women (https://www.redaccionmedica.com/secciones/industria/la-industria-farmaceutica-roza-la-paridad-directiva-y-la-supera-en-empleos-5861). Each scenario presented information about four hypothetical candidates (in the Spanish language). The attributes were the following:

i) “assessment by algorithm” with three levels: “sufficient”, “satisfactory” and “good”, indicating the degree of adjustment between candidate and job. This would be the most objective and reliable criterion.

ii) "co-worker’s recommendation” of the candidate with four levels: "I recommend this person totally", "this person causes an excellent impression", "this person does not inspire confidence", "I do not recommend this person", indicating the overall assessment of a human expert on the overall profile of the candidate (informed, but subjective criterion).

iii) “co-worker’s appraisal of the candidate profile picture” with three levels: “the profile picture inspires confidence in me”, “very attractive profile picture”, “not very professional profile picture” (the most subjective and unreliable criterion).

iv) "sex", with two levels: "female" and "male"; in each trial, sex was represented by symbolic images (see fig.1); the candidate’s sex was included as a control variable in the study in order to control a potential gender bias in the recruiter’s decision.

The levels of the four attributes that appeared with each candidate were randomized in each trial so that each of the candidates appeared with a combination of different levels. Figure 1 shows an example of a trial with a specific combination. With

1 In the original Spanish text, the levels for the hiring algorithm were “justo”, “acceptable”, “bueno”, indicating a low, medium and high level, in English. The copyeditor reminded us that the English translation (fair, acceptable, good) does not discriminate between levels in the same way as the Spanish original wording does. Therefore, we changed the words in this manuscript.
a click, participants had to indicate, in each of the 20 randomly created trials, which of the four hypothetical candidates they would choose for the next stage (a face-to-face job interview) given the combination of levels presented. To determine the minimum number of participants required, we previously carried out a design test using different sample sizes. A general guideline is to achieve standard errors (SE) of 0.05 or smaller for main effect utilities and 0.10 or smaller for interaction effects or alternative-specific effects. The test showed that for 20 trials and four attributes, a minimum of $N = 60$ was required for $SE \leq 0.05$. As we also intended to compare men and women, we set the sample size at 65 for each sex.

- Insert figure 1 approx here -

Questionnaire

To measure the participants’ self-perception in terms of gender roles, we used the short Spanish version of the Bem Sex Role Inventory (Zimmermann et al., 2011). The eight items of the instrumentality scale focus on achieving goals (for example, “convincing”), and the eight items of the expressiveness scale focus on the well-being of others (for example, "sensitive to the needs of others"). In the present study, the reliability of the subscale instrumentality was $a = .753$, and that of the subscale expressiveness was $a = .850$.

Procedure

Both the CBC task and the questionnaires were built using the Lighthouse Studio software (sawtooth.com). After explaining the project to each potential participant in a personalized message through LinkedIn, we sent them a link through
which we collected the data. Upon entering the application, they encountered the objective of the study and completed the informed consent process. In each step they received detailed instructions, indicating that they should select pharmacologists for a multinational laboratory. First, respondents answered the gender roles questionnaire; then they proceeded to the CBC task. Instructions for the CBC task were the following: “In each trial, you will choose a single candidate to be invited for a personal job interview. You must choose one and only one. To facilitate the task, previous profile assessments have been made.”

1. A hiring algorithm (automatic computerized program) has screened candidates according to objective criteria (studies, work experience, languages) from LinkedIn and a value is assigned to each candidate for the fit between the candidate and the job description: sufficient, satisfactory, or good.

2. A co-worker of yours (human resources professional) has appraised the candidate’s profile picture.

3. Another associate co-worker of yours (human resources professional) has studied the LinkedIn profile and given you a recommendation.”

The completion of both questionnaire and CBC task took the participants about 12-15 minutes.

Data analysis

Analysis of conjoint data yields a series of scores for each respondent at each attribute level (“part-worths”), as a measurement of the “utility” or value that the individual rater associates with a product (here: the candidate) and its attributes (Orme, 2014). These reflect the best estimate for the likelihood that the respondent will prefer
one choice (in this case, one of the four candidates) in comparison to a specified set of alternative choices. The higher the utility, the more preferred is the level; the sum of the utilities is equal to zero within each attribute. The relative importance of each attribute for the decision can be extracted using a hierarchical Bayesian analysis (HB) of the preferences of the respondents. The advantage of HB is that in the presence of sparse data (like in our case, where we used the minimum of participants statistically required), it improves modeling, because it provides estimates of individual part worths given only a few choices by each individual; it does so by taking information (means and covariances) from the data provided by the other respondents (Sawtooth Software, 2019). Other authors using CBC also use HB analysis in order to make their case stronger (Carbonell and Brand, 2018).

Contrarily to what happens with part-worths, these importance scores are ratio-scaled, and therefore parametric statistics can then be used to compare men and women and to calculate correlations with the scores obtained in the two questionnaire subscales; for the utilities, $\chi^2$-squares were calculated to assess the differences for each attribute.

Ethical clearance had been obtained from the first author’s university ethics committee. This work was supported by the Spanish Ministerio de Economía, Industria y Competitividad, MINECO) with a grant for Research and Development (reference: FEM 2016-76136-R). Declararions of interest: none

Results

Descriptive statistics of participants

Out of the recruiters who agreed to participate, 135 (67 women and 68 men) completed the questionnaire and all the CBC trials. These participants had a mean age
of 31.64 years (sd = 12.47), with an average work experience in recruitment of 5.36 years (sd = 5.42). There were no gender differences in relation to age ($t = -1.69$, $p = .866$, df = 133) or work experience ($t = -1.34$, $p = .162$, df = 133).

**Choice-based conjoint**

As could be expected, the participants conferred a significantly higher utility to the highest level of the attribute ("good") for screening by algorithm ($\chi^2 = 348.664$, df = 2, $p < .01$), to the highest level of the profile picture ("the profile picture inspires confidence in me") ($\chi^2 = 92.894$, df = 3, $p < .01$) and to the highest level of overall recommendation ("I totally recommend this person") ($\chi^2 = 1858.953$, df = 3, $p < .01$).

The sex of the candidates did not affect these utility values ($\chi^2 = 1.643$, df = 1, n.s.), i.e., there was no preference for one sex over the other. Regarding the combinations, there was no combined effect of sex and algorithm, sex and profile picture appraisal, sex and recommendation, or profile picture appraisal and recommendation. In other words, the participants valued equally in men and women maximum levels for algorithmic screening, profile picture and overall recommendation. However, there was an interaction effect of the combination algorithm and co-worker’s recommendation ($\chi^2 = 24.804$, df = 6, $p < .05$), and of the combination algorithm and appraisal of the profile picture ($\chi^2 = 14,256$, df = 6, $p < .05$). That is, the participants were willing to accept a lower algorithm score ("satisfactory") if the level of the recommendation was maximum ("I totally recommend this person"), with an average utility score of 0.508. They were also willing to accept a lower score for the algorithm, if the picture received the highest rating ("the profile picture inspires confidence"), with an average utility of 0.305. The Bayesian hierarchical analysis to establish the most important attribute for the participants at the time of making a decision confirms these data: the average of the
importance scores was highest for the co-worker’s recommendation (M = 58.69, sd = 9.47, followed by the algorithm (M = 22.34, sd = 11.23 and the profile picture (M = 15.84, sd = 6.49); the lowest importance was for the sex of the candidates (M = 3.11, sd = 3.76).

The participants’ sex and gender roles

Table 1 presents the descriptive statistics for the attribute importance values (in terms of zero-centred differences) and for the mean scores for instrumentality and expressiveness, separately for males and females, as well as Student’s t and intercorrelations.

- Insert table 1 approx here -

Women scored higher in expressiveness than men. No gender differences were found in relation to importance scores. In both male and female recruiters, the importance of the algorithm correlated negatively with the importance of the profile picture appraisal and the recommendation. For male recruiters, instrumentality correlated negatively with the importance of the candidate’s sex and positively with the importance of the co-worker’s recommendation. That is, men with high instrumentality gave less importance to the sex of the candidate on the one hand, and on the other, they relied more on the human expert. For female recruiters, the importance of the candidate’s sex correlated negatively with the algorithm and the co-worker’s recommendation, but there were no correlations between gender roles and importance scores. In women, gender roles were not associated with their preference for either sex.
Discussion

The main finding shows that recruiters of both sexes prefer the recommendations made by an expert person over the algorithm or other attributes, thus confirming hypothesis 1. This is even the case when the expert information is highly subjective (such as the co-worker’s rating of the profile picture “inspiring confidence”) and therefore should not be taken into account at all.

Neither the candidates’ nor the recruiters’ sex had an impact on the decision, and the recruiters’ gender had very little influence. This may be due to the researchers choosing a gender-neutral job profile as recruiters showed no preference for the candidate’s sex. Our results could have been different if the job description had been more gender-stereotypic, and thus, a possible gender-role congruity bias could have appeared. Recent studies show that gender stereotypes have changed (Eagly et al., 2020), but there is still a preference for recruiting males for male-dominated jobs (Koch et al., 2015). As for the participants’ characteristics, our results show that the recruiters’ own biological sex did not influence the importance of the candidates’ attributes. These results disconfirm with hypothesis 2, contradicting former studies in which men were more interested in using technology than women (Hargittai and Shaw, 2015). Nevertheless, the result does coincide with other studies, which found that despite men and women perceiving technology-related concepts differently, there were no gender differences with respect to performance (Sobieraj and Krämer, 2019). The participants’ gender had an influence only in the case of males, where instrumentality was related to lower importance values for the candidate’s sex and higher importance values for the co-worker’s recommendation, opposite to our expectations (disconfirmation of hypothesis 3). These inconclusive results might have to do with changes in gender stereotypes over the decades. On the one hand, traditional masculine and feminine roles
are losing importance, and on the other, women tend to assign themselves traits that had
been considered typically masculine, adopting a more androgynous profile (while men
do not the same with feminine traits (López-Saez et al., 2008; Oberst et al. 2016).

Another explication is the way gender roles were operationalized in our study: as
gender-related personality traits. A recent study (Hentschel et al., 2019) showed the
complexity of these constructs and concluded that studies should not just focus on
overall agency/masculinity and communality/femininity, but rather on their different
facets (such as leadership competence, assertiveness, concern for others, etc.). We might
wonder if it is possible that the instruments supposedly assessing gender roles actually
assess personality traits that are only loosely related to gender roles. Therefore, our
results on gender differences have to be interpreted with caution.

Our results are in line with other studies that show that, in general,
recommendations and descriptions of other people have a more significant impact on
the decisions of the user than objective aspects (Carbonell et al., 2012). Research shows
that expertise-based intuition, rooted in extensive experience within a specific domain,
plays a vital role in expert decision-making (Salas et al., 2010) and has a powerful
impact in decisions, even leading decision-makers to ignore important objective
information. Our results also concur with the previous literature on specific hiring
decisions. Indeed, hiring experts not only keep preferring intuitive decisions over
analytical procedures, but do so even if these are highly subjective (the candidate’s
profile picture) and should not be taken as significant. Earlier studies have tried to
explain this “stubborn” (Highhouse, 2008) tendency, by criticizing the companies’ and
professionals’ misconceptions about what are state of the art human resource practices.
Our findings add value to the existing literature and indicate that this stubbornness
prevails also over methods that use artificial intelligence. There is still little knowledge among professionals about what a hiring algorithm actually does, beyond allowing them to browse thousands of resumes in LinkedIn and other databases according to pre-established rules and criteria. This seems to be a general tendency for digitalized societies, with computers and algorithms mediating much of people's daily activity in one way or another (Dufva and Dufva, 2019). These authors believe that the consequences of digitalization are challenging to understand because most people lack first-hand experience.

Some authors predict that one-third of today’s jobs will be taken over by smart technology, AI, robotics and algorithms (STARA) by 2025 (Brougham and Haar, 2018). These authors show that greater STARA awareness has adverse effects on the employees’ commitment and career satisfaction. STARA awareness was also positively related to turnover intentions, cynicism, and depression.

**Practical implications**

As human resource management becomes increasingly digitalized, companies have to prepare the organization for the AI revolution, which also means rethinking the organizational structures and the roles of the managers (Institute for Employment Studies, 2018), for instance defining and refining the hiring professionals’ role in recruitment. At this point of the development, hiring algorithms are becoming indispensable to browse the large volume of online information about potential candidates, but perform mainly repetitive and “mundane” tasks (Upadhyay and Khandelwal, 2018), and are not meant to replace the human recruiter. Instead, hiring algorithms can simplify the work of the human recruiter by presenting a shortlist from which a final candidate can be chosen, but also improve the prediction of the future.
performance of the candidate, while the decision corresponds to the human professional. (Agrawal et al., 2019)

This function could leave recruiters with more time to interact with the candidates and to focus on strategic issues. HR issues are complex phenomena (Cappelli et al., 2018), and there are several challenges in using data science techniques in HR. Among these challenges, Cappelli et al. consider that it not at all clear what a “good employee” is, because complex jobs are interdependent with other jobs and therefore, individual performance cannot entirely be separated from group performance. Another challenge identified by these authors are concerns related to fairness and ethics. It might be difficult for a HR professional to explain to a candidate the AI-based criteria that led to a specific decision on the individual’s future, e.g. why they were rejected and somebody else with similar qualifications are not. Further, candidates and employees might try to outsmart the AI or show other adversarial reactions that could end up affecting the organizational outcomes.

Recruiters should be adequately informed about the actual functioning and potential (and shortcomings) of the technology used by their company, their potential worries about their tasks and the possibility of being replaced by the AI, have to be adequately addressed in order for recruiters to be willing to use digital hiring algorithms and rely on their outcomes. However, as with other professional activities, it is still to be seen how this will evolve in the future.

Limitations

Our experimental approach used a fictitious scenario in which participants could not make real choices with real consequences. In a real-life scenario, considerably more
variables influence hiring decisions. Research has shown that decision-making is influenced by a series of factors, especially the subjects’ decision-making style (Lodato, et al., 2011), their cognitive executive functions and their personality (Carbonell and Brand, 2018). Also, the participants were a small self-selected group among a large group of eligible professionals, and we did not distinguish between human resources professionals working inside a company or headhunters working externally.

Outlook and conclusions

In the context of highly sophisticated and supposedly more reliable selection methods using artificial intelligence, we may speculate wonder if there is another element that contributes to the professionals clenching to their own and others’ intuition and experience, despite the growing evidence of the AI’s efficiency. Two possible explanations come to mind. Recruiters may neglect or underestimate what a hiring algorithm based on AI actually does and therefore infer that human judgment is generally better. Alternatively, they might overestimate the potential of AI and be hesitant to use it because they fear it could replace them.

If recruiters underestimate the technology, they may tend to distrust and therefore underuse it or neglect its outcomes. On the one hand, hiring algorithms may be considered “dull” (doing just mechanical work), inflexible, or fallible. For example, inflexible parameters would might miss candidates with unusual career paths and thus deprive the company of high-potential talent. On the other hand, recent research (Caliskan et al., 2017) shows that adaptable machine learning technologies are not infallible; as they acquire human-like language abilities, they also absorb the implicit biases in language patterns, which may result in biased resume screening.
If recruiters overestimate the capabilities of hiring algorithms, they may also tend to distrust and underuse or disregard its outcomes because they fear the potential consequences of the frenzied developments in AI. Evidence was found (Elsbach and Stigliani, 2018) that people’s biases toward information technology involve its abstract and unseen characteristics, such as mysteriousness, non-humanness, and complexity and the concern that it may make decisions against human interests (technophobia, (Khasawneh, 2018). These biases and concerns result in beliefs about the technologies’ (un)trustworthiness. In this context, Brougham and Haar (2018) refer to the concept of “STARA awareness (STARA= smart technology, AI, robotics, and algorithms); they showed they showed that greater STARA awareness had adverse effects on the employees’ commitment and career satisfaction. STARA awareness was also positively related to turnover intentions, cynicism, and depression. This may occurs because users who do not possess technological expertise trust new technology vetted by others who are more experienced than they are. The consequences of the increasing digitalization are challenging for many people, because they lack first-hand experience of what digitalization actually feels like (Dufva and Dufva, 2019). As organizations invest more in these technologies to reduce costs and increase efficiency substituting workforce by capital (Agrawal et al., 2019), workers increasingly fear that this technology might replace them. This is not entirely improbable: in the next two decades, in the USA, 47 percent of jobs could be lost due to this effect, and in countries like China and India even more (Tobenkin, 2019), and 14 percent of the global workforce may need to switch occupational categories (Illanes et al., 2018).

In conclusion, our result of recruiters preferring to rely on human recommendation more than on AI technology in the field of recruitment may encourage further research on the
psychological correlates of the acceptance of AI technology in hiring, but also in other fields where AI is penetrating rapidly. Assuming that the AI revolution aims at substituting, supplementing and amplifying practically all tasks currently performed by humans (Makridakis, 2017), we suggest further studies in order to revise the current technology acceptance model in the context of STARA awareness (Brougham and Haar, 2018) and technophobia (Khasawneh, 2018). While doing so, it is important to take into account the potential users’ fear, misgivings or unease related to the technological developments in AI and its pervasive use. This unease could be attributed to the fear of losing one’s job in the near future, as presaged by scientists and echoed in the media or to the increasing awareness that algorithms and big data might come to know us better than we do.

References


Branscombe, N. R. and Smith, E. R. (1990), ‘Gender and racial stereotypes in


http://mc.manuscriptcentral.com/mrr
Institute of Employment Studies (IES) (2018), ‘Which way now for HR and organizational changes?’ Available at: https://www.employment-studies.co.uk/resource/which-way-now-hr-and-organisational-changes-ies-perspectives-hr-2018


Lewis, N. (2018), ‘Will AI remove hiring bias?’. Available at: https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/will-ai-remove-hiring-bias-hr-technology.aspx


Acknowledgements: The authors wish to thank Neil Daruwala for revising grammar and style of the final manuscript.
Tables and figures
Figure 1. Graphical outlay of a CBC task (example)
Table 1. Descriptive statistics, *t*-test and correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Males (N=68)</th>
<th>Females (N=67)</th>
<th>t (p) (df=133)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. candidate sex</td>
<td>2.56(3.20)</td>
<td>3.67(4.19)</td>
<td>-1.742(.084)</td>
<td>---</td>
<td>-.200(.-263*)</td>
<td>.071(.-307*)</td>
<td>-.125(-.380**)</td>
<td>-.257*(-.052)</td>
<td>-.166(-.059)</td>
</tr>
<tr>
<td>2. algorithm</td>
<td>22.89(12.21)</td>
<td>21.79(10.19)</td>
<td>.567(1.972)</td>
<td>---</td>
<td>-.575**(-.562**)</td>
<td>-.783**(-.607**)</td>
<td>-.134(-.183)</td>
<td>.077(-.059)</td>
<td></td>
</tr>
<tr>
<td>3. appraisal of profile picture</td>
<td>15.82(6.67)</td>
<td>15.87(6.34)</td>
<td>-.047(.962)</td>
<td>---</td>
<td>.012(-.209)</td>
<td>-.031(-.209)</td>
<td>.011(.157)</td>
<td>-.134(-.183)</td>
<td></td>
</tr>
<tr>
<td>4. recommendation by co-worker</td>
<td>58.73(9.88)</td>
<td>58.66(9.11)</td>
<td>.045(.964)</td>
<td>---</td>
<td>.270*(-.026)</td>
<td>-.049(-.221)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. instrumentality</td>
<td>41.36(5.47)</td>
<td>41.63(6.07)</td>
<td>-.261(.795)</td>
<td>---</td>
<td>.198(.043)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. expressiveness</td>
<td>41.72(7.33)</td>
<td>44.21(7.32)</td>
<td>-1.972(.05, 133)</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Correlations without parenthesis refer to male subgroup, those in parenthesis, to female subgroup; correlations with asterisks are significant on an alpha < .05 level.
Si estos fueran tus únicas opciones de un/a farmacéutico/a, ¿cuál escogerías?
(3 of 20)

<table>
<thead>
<tr>
<th>Sexo</th>
<th>Aceptable</th>
<th>Bueno</th>
<th>Justo</th>
<th>Justo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cría mediante algoritmo</td>
<td>“La foto no me gusta nada”</td>
<td>“La foto no me gusta nada”</td>
<td>“Foto muy atractiva”</td>
<td>“Foto no muy profesional”</td>
</tr>
<tr>
<td>Valoración de colaborador/a (foto)</td>
<td>“No recomiendo esta persona”</td>
<td>“Recomiendo esta persona totalmente”</td>
<td>“Esta persona no me inspira confianza”</td>
<td>“Esta persona da una impresión excelente”</td>
</tr>
<tr>
<td>Recomendación de colaborador/a</td>
<td>CBC2018_Random3 Select</td>
<td>CBC2018_Random3 Select</td>
<td>CBC2018_Random3 Select</td>
<td>CBC2018_Random3 Select</td>
</tr>
</tbody>
</table>
Table 1. Descriptive statistics, t-test and correlations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Males</th>
<th>Females</th>
<th>t (p)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. candidate sex</td>
<td>2.56</td>
<td>3.67(4.19)</td>
<td>-1.742</td>
<td>-.200</td>
<td>.071</td>
<td>-.125</td>
<td>-.257*</td>
<td>-.166</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.20)</td>
<td>(1.084)</td>
<td>(.084)</td>
<td>(.307*)</td>
<td>(.380**)</td>
<td>(.052)</td>
<td>(.059)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. algorithm</td>
<td>22.89</td>
<td>21.79</td>
<td>.567</td>
<td>---</td>
<td>-.575**</td>
<td>-.783**</td>
<td>-.134</td>
<td>.077</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.21)</td>
<td>(10.19)</td>
<td>(.572)</td>
<td>(.562**)</td>
<td>(.607**)</td>
<td>(.157)</td>
<td>(.183)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. appraisal of profile picture</td>
<td>15.82</td>
<td>15.87</td>
<td>-.047</td>
<td>---</td>
<td>.012</td>
<td>-.031</td>
<td>.011</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.67)</td>
<td>(6.34)</td>
<td>(.962)</td>
<td>(.209)</td>
<td>(.209)</td>
<td>(.015)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. recommendation by co-worker</td>
<td>58.73</td>
<td>58.66</td>
<td>.045</td>
<td>---</td>
<td>.270*</td>
<td>-.049</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.88)</td>
<td>(9.11)</td>
<td>(.964)</td>
<td>(-.026)</td>
<td>(.221)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. instrumentality</td>
<td>41.36</td>
<td>41.63</td>
<td>-.261</td>
<td>---</td>
<td>.198</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.47)</td>
<td>(6.07)</td>
<td>(.795)</td>
<td>(.043)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. expressiveness</td>
<td>41.72</td>
<td>44.21</td>
<td>-1.972</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(7.33)</td>
<td>(7.32)</td>
<td>(.05, 133)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Correlations without parenthesis refer to male subgroup, those in parenthesis, to female subgroup; correlations with asterisks are significant on an alpha < .05 level