Journal of Business Research 67 (2014) 1771–1779

Contents lists available at ScienceDirect

Journal of Business Research

Are two interviewers better than one?☆

Mario Fifić [a,b,*], Gerd Gigerenzer [b]

[a] Grand Valley State University, MI, USA
[b] Max Planck Institute for Human Development Center for Adaptive Behavior and Cognition, Lentzeallee 94, 14195 Berlin, Germany

ARTICLE INFO

Article history:
Received 1 April 2013
Received in revised form 1 November 2013
Accepted 1 January 2014
Available online 21 March 2014

Keywords:
Interview validity
Best interviewer
Wisdom of the crowd
Less is more
Free riding

ABSTRACT

How many interviewers per job applicant are necessary for a company to achieve the highest hit rate? Are two better than one? Condorcet's Jury Theorem and the "wisdom of the crowd" suggest that more is better. Under quite general conditions this study shows, surprisingly, that two interviewers are on average not superior to the best interviewer. Adding further interviewers will also not increase the expected collective hit rate when interviewers are homogeneous (i.e., their hits are nested), only doing so when interviewers are heterogeneous (i.e., their hits are not nested). The current study shows how these results depend on the number of interviewers, their expertise, and the chance of free riding, and specify the conditions when "less is more". This analysis suggests that the best policy is to invest resources into improving the quality of the best interviewer rather than distribute these to improve the quality of many interviewers.

Introduction

When consulting firms hire candidates as business consultants, or university departments invite applicants for faculty positions, the final decision is often based on a series of interviews. How many interviewers should be used for each candidate to achieve the best results? At first glance, the answer seems to be: the more, the better. For instance, the Condorcet's Jury Theorem says that the probability of a correct decision between two options increases with the number of decision makers in the group, provided that the individual probabilities of a correct decision are all greater than chance (Condorcet, 1785). Galton's (1907) seminal work on the vox populi appears to suggest the same conclusion, as does Bernoulli's law of large numbers. Modern concepts such as swarm intelligence (Krause & Ruxton, 2002) have led to speculations that if a diverse group can outperform an expert, then even CEOs might be in less demand in the future (Surowiecki, 2004). Do these arguments apply to interviewers as well?

The research reported in this article was motivated by a period in which one of us advised a consulting firm on their recruitment process. The firm has some 10,000 applications per year from young aspirants for over 100 open positions. Its decision-making process was neither fast nor frugal. In a first round, all applicants were evaluated on the basis of their CVs, statements, and letters, and about 500 were selected. In a second round, these selected applicants were flown in, put up in the best five-star hotel in town, and grilled by three interviewers, after which about half of them were eliminated. In a third round, a few weeks later, the remaining applicants were flown in again, put up in elegant suites, and quizzed by three other interviewers. For the final choice, the interviewers met to vote; offers were made to those with the highest number of votes. Millions of dollars were spent on the direct and indirect costs of this process even though the firm had no systematic quality control and kept no electronic records until a few years ago.

Companies around the world depend on interviews as a tool for selecting the best candidates. The consulting firm above represents a typical (but not isolated) case in which the question about the best number of interviewers was never considered.

The validity of an interview is typically defined as how effective a certain method is in finding the best candidates. The validity can be quantified by keeping records of the hired candidates' advances on the corporate ladder. Several meta-analyses have been conducted on a large body of published studies showing that the interview validity coefficient can vary from low-end values of .10 (Dunnette, 1972) and .22 (Hunter & Hunter, 1984) to moderate values of .3 to .6 (Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Wiesner & Conshaw, 1988). These numbers are correlation coefficients between the interview test outcomes and criteria — measures of professional success. Meta-analyses show that improving interview validity is possible by controlling for various factors: Among the most important are the amount of structure imposed during the interview (Huffcutt & Arthur, 1994; Schmidt & Zimmerman, 2004), the interviewer selection and training (Conway, Jako, & Goodman, 1995; Huffcutt & Woehr, 1999),

☆ The authors are grateful for the comments and insights by the members of Adaptive Behavior and Cognition (ABC) group, Max Planck, Berlin, for improving earlier versions of this paper.

* Corresponding author at: Grand Valley State University, Psychology Department, One Campus Drive, Allendale, MI 49401, USA.
E-mail addresses: fificm@gvsu.edu (M. Fifić), gigerenzer@mpib-berlin.mpg.de (G. Gigerenzer).

http://dx.doi.org/10.1016/j.jbusres.2014.03.003
0148-2963/

and the method of aggregation of independent reviewers' decisions (Dreher, Ash, & Hancock, 1988). Some meta-analyses indicate the significant impact of the number of interviewers on interview validity. Interview validity appears to improve as the number of interviewers increases (Conway et al., 1995; McDaniel et al., 1994; Schmidt & Zimmerman, 2004; Wiesner & Conshaw, 1988). Thus, in recruitment practice these findings appear to confirm Condorcet's Jury Theorem.

How to choose the number of interviewers?

The observations at the consulting firm led us to ask: Are more interviewers always better? Is there a systematic way to relate the number of interviewers to the resulting quality of the hiring process? The current study focuses here on the question of how many interviewers are needed to select the best m candidates out of a pool of size M, and exclude other goals that are simultaneously pursued in actual recruiting, such as to impress a candidate by an elaborate selection process, or to familiarize the faculty with the candidates. Before answering the question, the authors of the current study first checked whether such an elaborate process is typical in consulting firms. The authors retrieved information on a sample of companies, including 3M, Bain & Co, Booz & Co, Boston Consulting Group, Deloitte, Cargill, McKinsey & Co, PriceWaterhouseCoopers, Thomson, and Thrivent Financial for Lutherans. The number of interview rounds varied between 2 and 5 (on campus or in the company office), and the number of different interviewers per candidate varied between 5 and 11, depending on position (e.g., associate or senior consultant) and company. This informal survey revealed that an elaborate step-wise process is not uncommon, and that multiple interviewers appear to be standard.

A body of research compares collective decision making and individual experts' opinion. Specifically, this research has addressed the effect of how to combine individual votes into a collective vote, from various forms of aggregation such as the majority rule (Arkes, 2003; Hastie & Kameda, 2005; Reimer & Katsikopoulos, 2004; Sorkin, West, & Robinson, 1998) to averaging of individual judgments (Ariely & Levav, 2000; Armstrong, 2001; Clemen, 1989; Clemen & Winkler, 1987; Einhorn, Hogarth, & Klempner, 1977; Gordon, 1924; Hogarth, 1978; Johnson, Budescu, & Wallsten, 2001; Wallsten, Budescu, Erev, & Diederich, 1997; Winkler & Poses, 1993). One conclusion drawn is that more experts do better, consistent with Condorcet's Jury Theorem. Various amendments have been reported, such as that the increase is inversely related to the average inter-correlation among individual experts' opinions (Hogarth, 1978). In other words, adding new experts to an existing group leads to little improvement if the new experts make similar decisions (e.g., Winkler & Clemen, 2004). A second conclusion is that more is not always better. Several researchers note that the best experts in a group sometimes outperform the group's collective score (e.g., Gordon, 1924). In a study of physicians' performance in an intensive care unit (Winkler & Poses, 1993), the best prediction of patients' survival rates was obtained by taking averages of performance in the two best individually performing groups in the hospital rather than in all groups. Likewise, in a study of economists' ability to predict economic growth, the forecasts of economists with the best previous histories of performance were better than a combined group score (Graham, 1996).

Several other studies focusing on the individual measures of interviewer validity show that some interviewers are better than the others in selecting the best candidates (e.g., Dipboye, Gaugler, Hayes, & Parker, 2001; Ghiselli, 1966; Heneman, 1975; Pulakos, Schmitt, Whitney, & Smith, 1996; Yonge, 1956; Zedeck, Tziner, & Middlestadt, 1983). The implication of these studies is that adding more interviewers might harm the selection personnel process, thus potentially implying the contrary to Condorcet's Jury Theorem.

The existence of free riders is a proposed resolution of this apparent contradiction—the phenomenon that with increasing group size, the extent of experts' involvement in the group decreases (Albanese & van Fleet, 1985; Kameda, Tsukasaki, Hastie, & Berg, 2011; Kerr & Tindale, 2004). In consequence, the quality of collective decision making may decline. As the team grows larger, individual experts tend to feel less responsible for collective decision making and invest less in information accrual. Free riding is predicated on the belief that someone else in the team will collect and process the relevant pieces of information. The evidence for free riding has been investigated in criminal law for determining the right jury size, not too big and not too small (Mukhopadhaya, 2003). In organizational economics, some researchers have been argued that larger groups lead individual members to engage less in information acquisition (Holmstrom, 1982). In social psychology, free riding is attributed to individuals' loss of motivation to contribute to social groups (Kerr & Tindale, 2004).

We aim here at a more general analysis of the conditions under which "less is more" in choosing the right number of interviewers, including interviewer characteristics and the free riding phenomenon.

Setting and terminology

In this article, the authors derive a systematic answer to the question of how many interviewers are needed to select the best candidates. To do so, one first needs to define the setting, which the authors model after the situation in many large consulting firms, as described above. The task is to pick the m best candidates out of a pool of size M. The m top candidates are called *targets*. All other candidates are called *non-targets*. Each interviewer i is characterized by a *hit rate* h_i, which the authors define as the relative frequency of correct target identifications among the interviewer's m votes. A hit rate h_i defined here could be interchangeably used with the term *interviewer's selection validity*, as both can be used to measure the efficiency of personnel interview to predict future job performance of hired candidates. For instance, if $m = 10$, a hit rate of .8 means that an interviewer has an expectation (or long-run frequency) of correctly identifying 8 out of the 10 targets, while missing two and voting for two non-targets (false positives). In this setting, interviewers differ in h_i, and the identity of the best interviewer is known (e.g. Dougherty, Ebert, & Callender, 1986; Ghiselli, 1966; Yonge, 1956). In addition, pairs of interviewers can differ in *homogeneity* in judgment (defined below), which reflects the kind of cues they look for and the strategy for processing these cues.

Each interviewer conducts the interview alone and independently votes yes/no for each candidate to be hired (*interviewer independence*), with the constraint that the number of yes-votes equals m. Finally, the votes of the N interviewers are added up to determine who survives to the next round or who will be made an offer (this is called the *majority rule*, as in Condorcet's Jury Theorem). The majority rule specifies that each vote counts equally and the group decision is the tally of votes (Hastie & Kameda, 2005). In case of a tie between candidates, offers will be decided randomly. The ties are candidates who received an equal number of votes, but of whom only a subset can be selected as top m candidates. The resulting hit rate of the N interviewers achieved by applying the majority rule is their *collective hit rate*.

A team of N interviewers can be either *homogeneous* or *heterogeneous*. Consider the case of two interviewers. They form a homogeneous (nested) set if and only if the second interviewer's correct identifications form a subset of those chosen by the first interviewer. A team of homogeneous interviewers is likely if everyone has been trained to use similar cues to identify top candidates. Two interviewers form a heterogeneous team if their correct identifications are not nested. If two interviewers are heterogeneous, they are likely to rely on different cues to identify the best candidates. A heterogeneous team could be formed with the purpose of covering a broad range of interviewer experience using a large range of cues. Such interviewers will complement each other, focusing on identification of cues that fall outside the other's domain of expertise.

We now turn to the main question. Compared to what the best interviewer can achieve alone, does adding more interviewers lead to better results? Let's begin with the simplest case of two interviewers and then