

Disregarding Preliminary Information When Rating Job Applicants' Performance: Mission Impossible?¹

TALYA MIRON-SHATZ²

*Center for Health and Wellbeing
Princeton University*

GERSHON BEN-SHAKHAR

Hebrew University

Firms allocate considerable resources to test job applicants' skills and to ensure that hiring decisions are nondiscriminatory. Interpreting selection output in light of preliminary information may undermine the impartiality of personnel selection decisions. In this study, human resource managers were presented preliminary information about a candidate's performance and asked to rate him according to a detailed assessment-center report of his performance. We examined several interventions for reducing reliance on preliminary information: retrieving assessment-center information, generating a rating model, or both. Participants ($N = 167$) excluded preliminary information from the interventions, but relied on it when rating the candidate. The documented failure to control for the effects of preliminary information on subsequent judgments can contribute to improved selection procedures.

Firms approach the designing of personnel selection tools and procedures with great scrutiny, aiming to avoid discriminatory practices. This scrutiny may be undermined if human resource (HR) managers interpret selection output in light of preliminary information regarding the applicant.

The present paper tackles the processing of relevant information, such as a candidate's accomplishments, which might prove to be an even greater challenge for professionals. These professionals need to evaluate candidates uniformly, based on the selection tools the organization applies. Indeed, questions are raised regarding the measurement of predictors and performance criteria (Saad & Sackett, 2002), and the tacit implementation of selection rules (Lewis & Sherman, 2003). On top of these concerns, organizations

¹The research reported in this article was made possible by grants from the Israel Foundation Trustees (2002–2004); the Levi Eshkol Institute for Social, Economic, and Political Research in Israel; the Minerva Center for Human Rights; the Hebrew University Faculty of Law; and the Harvey L. Silbert Center for Israel Studies. The authors are grateful for the helpful comments of Maya Bar-Hillel, Gideon Keren, Amnon Rapoport, Ilana Ritov, and Yaacov Schul. Segments of this work were presented at the 19th Subjective Probability, Utility, and Decision Making conference; the 9th Behavioral Decision Research in Management conference; Fuqua Business School, Duke University; and the 25th Society for Judgment and Decision Making conference. The article is based on the doctoral dissertation of the first author.

²Correspondence concerning this article should be addressed to Talya Miron-Shatz, 314 Wallace Hall, Center for Health and Wellbeing, Princeton University, Princeton, NJ 08544-1013. E-mail: tmiron@princeton.edu

are exposed to litigation challenging selection devices as being discriminatory (Terpstra & Kethley, 2002).

In the present study, we gave HR managers preliminary information about an applicant's overall performance, but asked them to rate only his assessment-center performance, of which they read a detailed report. We also examined three interventions for reducing reliance on preliminary information: retrieving positive and negative information about assessment-center performance, generating a model for rating his assessment-center performance, or both. The success of the interventions hinged on two sets of variables, measuring the presence of the preliminary information in participants' execution of the interventions and in their final ratings of the applicants' assessment-center performance.

Reliance on Preliminary Information

That one's initial hypothesis or impression influences one's final judgment is an established phenomenon (e.g., Chapman & Elstein, 2000; Moser, Gadenne, & Schröder, 1988). Relevant preliminary information (e.g., a job applicant's curriculum vitae, letters of recommendation) abounds in the personnel selection scene and can form the basis for an initial hypothesis that would impair subsequent judgments of selection material (Macan & Dipboye, 1990; Philips & Dipboye, 1989). To illustrate, recruiters' interpretations of candidates' interview performance were biased toward the pre-interview impression formed by the application letter (Macan & Dipboye, 1994).

A number of mechanisms sustain this agreement effect of forming and confirming a preliminary hypothesis that feeds on preliminary information (Koehler, 1993): (a) difficulty in retrieving information that conflicts with the initial hypothesis or, if retrieved, a tendency to ignore it; (b) assignment of relatively high weights and high (subjective) reliability scores to evidence support of the initial hypothesis; and (c) outcome bias. Research has found expert judgments to be plagued with a reliance on preliminary information, even when this information is of a questionable nature. For example, a brief, anonymous diagnostic hypothesis attached to a patient's file, containing psychological interviews and projective tests influences the diagnosis (Ben-Shakhar, Bar-Hillel, Bilu, & Sheffer, 1998).

Experts have tended to rely on preliminary information when exercising judgment in their home ground, whether that information is gleaned from clinical psychology (Chapman & Chapman, 1967, 1969, 1982; Oskamp, 1965), science (Koehler, 1993), or accounting (Cloyd & Spilker, 1999, 2000). Given the widespread reliance on preliminary information and the sensitivity of personnel selection decisions, the present study examines the efficacy of

instructions to disregard preliminary information and interventions for reducing its effect on experts, who regularly make hiring decisions. If information is both prevalent and potentially disruptive, the obvious cure would be to ignore it. Yet, the extant literature indicates that ignoring and disregarding are rather hard to accomplish.

Several factors facilitate disregarding. Disregarding of relevant information depends on the degree to which the information to be disregarded has been assimilated in the target person's mental representation and on the ease with which the implications of the preliminary information could be disconnected from the representation (Macrae, Bodenhausen, & Milne, 1995). Such singling out is harder the more the information to be disregarded resembles the "legitimate" information (Burnstein & Schul, 1982). Participants distinguish between relevant and nonrelevant information and use these pieces of preliminary information differently: They make inferences involving discredited information when it affords causal structure, but manage to disregard it when its introduction seems random (Johnson & Seifert, 1994).

The cover story we used was designed to minimize the relevance of the preliminary information to the judgment task. We told participants that the study was designed to examine the unique contribution of each segment of the selection portfolio. This has ecological validity, as the participants were professionals who constantly engage in personnel selection, so they are aware of the costs of using each selection tool and the ensuing organizational debates regarding the necessity of each tool.

We offered a reward as an incentive to base the judgments only on the assessment-center summary. This encouraged participants to disregard the preliminary information, despite its relevance to the "legitimate" assessment-center summary that they were about to rate. This also increased the study's ecological validity because, in reality, preliminary information about candidates is often available to HR managers, who are supposed to apply selection procedures independent of this information.

Still, the preliminary information to be disregarded refers to the candidate's selection file and could have suggested a causal structure for interpreting the candidate's assessment-center summary, which was included in the file. This setting renders it difficult to disconnect the implications of the material to be disregarded from the subsequent judgment, regardless of the specific demand to do so. To this effect, we applied several interventions to reduce reliance on preliminary information.

Interventions to Reduce Reliance on Preliminary Information

Interventions to reduce, if not eliminate, reliance on preliminary information tackle the "mismatch between judgment and task" (Fischhoff, 1982, p. 422)

by forcing participants to express their knowledge and to seek conflicting data. Reliance on preliminary information is attenuated by procedural methods such as having participants choose among available pieces of information and justify their choice, rather than justify their final decision (Jonas, Frey, & Greenberg, 2003; Jonas, Schulz-Hardt, Frey, & Thelen, 2001). However, raising participants' motivation to judge accurately often fails to reduce reliance on preliminary information (Evans, Newstead, Allen, & Pollard, 1994).

We used structure-modifying interventions, which presumably alter the task and disrupt the tendency to proceed in accordance with the preliminary information. Structure modification through urging participants to consider an alternative hypothesis has been shown to be effective in reducing reliance on preliminary information (Hirt & Markman, 1995; Koehler, 1991). In addition, our interventions involved acquiring, or at least applying, a decision skill, such that we thought might generalize beyond its immediate application to future selection decisions.

Structured recall: Improving information availability. The assumption that memory leads to judgment, rather than vice versa is somewhat naïve (Hastie & Park, 1986). Selectivity in encoding or retrieval stems from the fact that once the judgment is made, a search for supportive evidence is conducted. This is a two-stage process in which a quick judgment is formed, followed by more systematic (yet slanted) processing.

Interventions during the data-collection phase usually require collection of evidence both supportive and disconfirmatory of the preliminary judgment, thereby minimizing such selectivity and increasing rating accuracy (Woehr & Feldman, 1993; Woehr & Huffcutt, 1994). Our intervention promoted the retrieval of positive and negative evidence about the candidate (which could either support the preliminary information or contradict it) in varying order (Baltes & Parker, 2000; Koriat, Lichtenstein, & Fischhoff, 1980).

Generating a decision model: Formalizing the judgment process. The literature is ripe with evidence regarding the superior predictive validity of model versus clinically based judgments (Meehl, 1954; Sawyer, 1966). Our participants generated models comprised of criteria and their respective weights for rating the candidate's assessment-center report. Consumer research has indicated that model building occurs spontaneously and without formal training (Murthi & Srinivasan, 1999; Shugan, 1980).

Participants could benefit from this intervention in two ways. The first benefit, as suggested by the bootstrapping literature (Dawes, 1979), is that consistently applied random weights can outperform expert judgment. Participants' ratings of the candidate's assessment-center summary could thus be more valid than if comprised otherwise. The second benefit of the model is

that participants had to read and evaluate the information regarding the candidate's performance in a comprehensive manner in order to assign weights. They did this instead of focusing on the aspects of the candidate's performance that match the preliminary information to be disregarded. Our search of the literature indicates that generating a decision model has not been used previously as an intervention for formalizing the judgment task and reducing reliance on preliminary information.

Combined interventions: Structured recall and generating a decision model. We found but one previous study that required participants to undergo two interventions. This endeavor was conducted in a somewhat related field; namely, reducing memory errors as a result of a shift between the standard to which a target is compared during judgment and retrieval. While each intervention was effective, combining them was not, and the effect was attributed to cognitive overload (Higgins & Liberman, 1994). Our mission expanded that of Higgins and Liberman, both by using other interventions and by examining whether these findings could be replicated with experts.

If these interventions were effective, it would be worthwhile examining their carryover to similar judgment tasks, taking place 2 weeks later. A prolonged effect would indicate that the interventions had made a dent in the experts' judgment repertoire.

HR Managers as Experts

This study is designed to estimate the extent of HR managers' reliance on preliminary information and the efficacy of various interventions on this particular population, which we refer to as *experts*. Some definitions of expertise (Ganzach, 1994; Showers & Cantor, 1985) focus on improved thought processes, for which we did not test our participants. Similarly, we did not pre-select them according to the recommendations of colleagues (Shanteau, 1988). Such pre-selection could have reduced the generalizability of our findings. Instead, we adopted a pragmatic stance, considering someone who acquired his training and experience at a certain profession or occupation an expert (e.g., Schulz-Hardt, Frey, Luthgens, & Moscovici, 2000). Hence, our expert participants were HR managers who regularly make personnel selection decisions, the quality of which remains unknown to us, along with the nature of the training and feedback they receive.

Goals of the Present Study

The initial goal of this study is to examine whether experts will follow instructions to disregard preliminary information so that it will not affect

their final ratings of an applicant's performance. As the literature points to the difficulties in disregarding, we also examine several interventions for reducing reliance on preliminary information. An effective intervention will be performed without reference to preliminary information, and the applicant's rating associated with it will not vary according to the valence of the preliminary information. With the design of effective interventions in mind, an important goal is to see whether the combined intervention will yield excessive reliance on preliminary information and whether the interventions will have a prolonged effect.

Research Hypotheses

The following hypotheses are proposed for the present study:

Hypothesis 1. Participants will disregard preliminary information when rating the candidates so that the ratings in the control condition will not vary according to whether the preliminary information was positive or negative.

Hypothesis 2. Participants will not relate to preliminary information when executing the interventions.

The valence of the retrieved information will be the same under positive and negative preliminary information conditions, as will the valence of the decision models.

Hypothesis 3. Participants who perform the interventions will rely less on preliminary information when rating the candidate's assessment-center performance than will participants in the control group.

Operationally, the difference in judgments observed under the positive and negative preliminary information conditions will be lower in the intervention conditions, relative to the no-intervention condition (i.e., control group).

Hypothesis 4. The combined intervention will be associated with greater reliance on preliminary information, compared with the separate interventions.

Method

Participants

HR managers ($N = 167$; 112 women, 55 men) participated in a two-session experiment with sessions scheduled 2 weeks apart. Participants' mean age

Table 1

Candidate Ratings in First Measurement by Experimental Group

	Positive preliminary information			Negative preliminary information			Effect size
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
Control group	6.72	1.33	30	5.80	1.70	28	.60
Model generating	6.74	1.32	19	5.72	2.00	16	.61
Structured recall (R)	7.42	1.43	18	6.63	1.38	23	.55
Combined (R + B)	7.06	1.29	17	6.44	1.36	16	.48
Total	6.94	1.35	84	6.14	1.64	83	.53

Note. Ratings were made on a 10-point scale ranging from 1 to 10.

was 41.1 years ($SD = 10.8$), and they had an average of 6.25 years of relevant experience ($SD = 5.86$). We recruited in HR management seminars and through mailing lists.

Design

We used a 2×4 between-subjects factorial design in which the valence of the preliminary information (positive or negative) to be disregarded and the type of intervention (structured recall, decision model, combination of both, or none) were manipulated independently. Participants were assigned randomly to the eight experimental conditions. Each intervention cell had approximately 20 participants, and each control cell had approximately 30 participants (exact numbers are displayed in Table 1).

Stimulus Material

Two assessment-center summaries (one for each applicant) were written specifically for this research paradigm. Each summary related to an applicant's performance along the following categories: leadership, organizational and managerial skills, communication skills, reaction to pressure, cognitive abilities, personality resources, and interpersonal skills. Categories were comprised of several germane items. For example, the category of interpersonal skills was comprised of items such as "The applicant forms warm and trusting relationships with others."

Manipulations

All of the participants received *instructions to disregard* the preliminary information. We told them that the study examined the specific contribution of each segment of the applicant's file (unlike the preliminary information, which referred to his overall performance). After reading the applicant's assessment-center summary, participants estimated how people who only read the assessment-center summary (and who did not receive the preliminary information, a point that was emphasized repeatedly) rated the applicant. We drew a \$250 reward from among the participants whose ratings were the closest to the average rating of the raters who only read the assessment-center summary (the average ratings of the control groups served as the target score). We manipulated two variables: valence of the preliminary information about the candidate, and the intervention for reducing reliance on the preliminary information.

Preliminary information. Participants' initial impressions of the applicant were manipulated by telling them that, having read the applicant's complete file (containing a letter of application, the summary of a structured interview, the results of a computerized test of cognitive abilities and personality traits, and the assessment-center summary), the director of the supermarket chain thought that the chain should [should not] hire the applicant. We quoted the director as saying "The applicant is in the top [bottom] quarter of the candidates. He is a very good [bad] applicant" (wording adapted from Baltes & Parker, 2000, p. 247).

Interventions for reducing reliance on preliminary information. We applied the following interventions (as well as the distracter task) in the first experimental session, so that the following session would measure their prolonged effect. Participants in the structured recall intervention condition completed the following sequence: They received 3 min to write "all the good things" they remembered about the candidate, and another 3 min to write "all the bad things" about the same candidate. The order was counterbalanced across participants. This was an abbreviated version of Baltes and Parker's (2000) structured recall intervention, as their participants watched a 15-min video interview, while our participants read a one-page assessment-center summary.

Participants in the decision model intervention condition completed the following sequence: They read a brief explanation of a decision model and an example from a different context.³ They then generated a decision model,

³The explanation was as follows: "Every person chooses cakes differently. Some consider flavor to be the most important factor and price to be of moderate importance. For such a person, the effect of flavor on the decision will be relatively large (e.g., flavor will be assigned a 50% weight), and price will have a lesser effect (e.g., it will be assigned a 25% weight). The remaining 25% will be divided between all other factors."

comprised of whichever criteria they chose, along with their respective weights. The model was meant to form the basis for their estimation of the average ratings the judges gave to the candidate's assessment-center summary. There were eight empty lines for criteria and weights, and participants were not instructed to use the seven assessment-center categories as criteria.

Participants in the combined intervention condition first completed the structured memory intervention. Then, they completed the decision model intervention.

Participants in the no-intervention condition (i.e., control group) answered a few questions about their occupations and specified where they were born and where they live. The questions were pre-tested to last approximately as long as the separate interventions (the combined intervention inevitably took longer).

Procedure

In both experimental sessions, we gave participants a booklet with all of the required information. First, they read a cover story containing the preliminary information about the candidate (positive or negative) and received the instructions to disregard it in their final rating, because the study examined the unique contribution of each component of the selection portfolio.

After reading a summary of the candidate's performance in an assessment center, participants in the intervention conditions went through the interventions, whereas those in the control condition completed the distracter task. Subsequently, all participants estimated the average rating of the candidate's assessment-center performance by others who have only read that segment of the selection file, using a rating scale (described later). Participants then specified what influenced their ratings.

We repeated this procedure 2 weeks later, with each participant evaluating a new applicant (having randomly divided the order between participants). The intervention or distracter task was not included in the second measurement, however. Finally, we e-mailed the participants a short debriefing, thanked them, and enclosed a coupon to a book chain for their participation.

Dependent Variables

Participants estimated the average rating given to a candidate's assessment-center summary by judges who read only this summary. The rating was made on a 10-point scale ranging from 1 (*very bad*) to 10 (*excellent*).

There are two other dependent variables that measure the extent to which the preliminary information was reflected in participants' execution of the interventions. These measures were based on independent experts' valence ratings of the evidence included in the assessment-center summaries:⁴

1. For every participant in the information-retrieval condition, we calculated total retrieval valence as the mean of the valences across all items that he or she retrieved (under the headings of *good* and *bad*) about a candidate.
2. For every participant in the decision-model condition, we computed the decision-model valence by multiplying the valence of each category included in the model, as determined by experts' ratings with the relative weight assigned to it by the participant.⁵

We used a rejection region of $p < .05$ for the statistical tests, unless otherwise indicated. Effect-size estimates were computed using Cohen's (1988) f values.

Results

Means and standard deviations of the ratings given to applicants' assessment-center summaries in the first measurement, as well as effect-size estimates computed for the intervention conditions and the control group are presented in Table 1. The results displayed in Table 1 were subjected to a 2×4 between-subjects ANOVA, with valence of preliminary information (positive vs. negative) to be disregarded and intervention (structured recall,

⁴We asked four occupational psychologists who specialize in personnel selection (M experience = 8.38 years), and who did not participate in the core study, to rate each of the 66 items that appeared in the assessment-center summaries. They rated the items on a 10-point scale ranging from 1 (*very bad*) to 10 (*excellent*). Each item described a particular characteristic or behavior of a job candidate (e.g., "Delivers his messages clearly and moderately"). Item valence was determined by the mean of the expert ratings it received. Mean inter-judge reliability (Spearman rank-order correlation) was .623. Reliability of the composite score by the Spearman-Brown formula was .868. Items were grouped into seven categories (e.g., organizational skills, managerial skills). The valence of a category was determined by mean expert valence ratings across all items in that category.

⁵For example, the 100th participant included the following categories in his model: leadership (applicant rating = 5.50; 23%), organizational and managerial skills (rating = 7.13; 20%), reaction to pressure (rating = 4.88; 5%), communication skills (rating = 6.63; 5%), cognitive abilities (rating = 6.60; 10%), personality resources (rating = 6.57; 20%), and interpersonal skills (rating = 7.33; 17%). Hence, the overall score of the applicant's assessment-center summary according to the model was 6.49.

decision model, combination of both, or none) as independent variables. This analysis reveals a statistically significant large effect for the factor of valence of the preliminary information to be disregarded, $F(1, 159) = 12.38$ ($MSE = 2.23, f = .27$), indicating that participants who received positive preliminary information about the candidates rated them more favorably ($M = 6.94, SD = 1.35$) than participants who received negative preliminary information ($M = 6.14, SD = 1.64$), despite instructions to ignore the preliminary information.

An inspection of the effect sizes displayed in Table 1 indicates that overall, the interventions were associated with an effect size similar to that of the control group. A planned comparison analysis contrasting the effect size in the control group with the effect sizes observed under all of the intervention conditions (decision model, structured recall, and combined) was not statistically significant, $t(167) = 0.37, ns$.

We conducted another planned comparison to examine the influence of the preliminary information to be disregarded on the final ratings in the intervention groups. The effect sizes for the decision model and structured recall conditions ($ds = .61$ and $.55$, respectively) were larger than for the combined intervention condition ($d = .48$). This difference was not statistically significant, $t(112) = 0.46, ns$. Participants' experience (i.e., number of years they have been engaged in personnel selection) had no effect on their ratings of the candidates.

A $2 \times 4 \times 2$ between-subjects ANOVA, with experience (below and above the median), valence of the preliminary information (positive vs. negative) to be disregarded, and intervention (structured recall, decision model, combination of both, or none) as independent variables does not reveal statistically significant effects for experience, $F(1, 150) = 0.39$ ($MSE = 2.21, f = .01$), or for its interactions with the other factors. Similar ANOVAs do not reveal statistically significant effects for age, $F(1, 150) = 0.01$ ($MSE = 2.44$), gender, $F(1, 150) = 0.00$ ($MSE = 2.47$), or for their interactions with the preliminary information and intervention.

Effect of Preliminary Information on Valence of Retrieved Items and Decision Models

Apart from examining the effect of the instructions to disregard on the final ratings, we looked at their impact on the degree to which the information retrieval and the decision model valence scores were affected by valence of the preliminary information. Means and standard deviations of the valences of the retrieved information and the decision models as well as effect size estimates are presented in Table 2. ANOVAs conducted on the mean

valences of the retrieved information and the decision models reveal that neither was affected by the valence of the preliminary information.

Specifically, a 2×2 between-subjects ANOVA conducted on the retrieval score with valence of the preliminary information about the candidate (positive vs. negative) and intervention (information retrieval vs. the combined intervention) as dependent variables did not produce any statistically significant effects, $F(1, 68) = 0.36$, *ns*, for valence of the preliminary information, and $F(3, 68) = 0.05$, *ns*, for its interaction with the interventions ($MSE = 0.38$, $f = .06$ and $.08$, respectively). A similar 2×2 between-subjects ANOVA conducted on the valence of the decision model did not produce statistically significant outcomes, either for the main effect of the preliminary information, $F(1, 68) = 1.51$ ($MSE = 1.95$, $f = .14$), or for its interaction with the interventions, $F(3, 68) = 0.13$ ($f = .04$). This indicates that participants followed the instructions to disregard when they performed the interventions, and there is no evidence that they were affected by the preliminary information.

An inspection of Table 2 reveals that all effect sizes were considerably smaller than those obtained with the ratings of the assessment-centers'

Table 2

Total Valence Ratings of Interventions

	Positive preliminary information			Negative preliminary information			Effect size
	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	
Structured recall	5.57	0.49	17	5.51	0.50	22	.13
Combined	5.47	0.58	17	5.46	0.52	16	.01
Overall retrieval	5.52	0.53	34	5.49	0.50	38	.05
Model generating	5.57	0.81	17	5.74	0.63	14	-.16
Combined	5.71	0.77	17	5.42	1.57	16	.24
Overall model	5.64	0.78	34	5.57	1.22	30	.06

Note. Valence ratings are based on independent experts' ratings of the items that appeared in the assessment-center reports. Model ratings were calculated by multiplying experts' valence ratings of categories of items by the relative weights assigned to them by participants. Memory ratings were calculated by averaging the independent experts' valence ratings across all retrieved items. There were 3 participants who did not write down the recalled information, and 1 participant did not write down the model, hence the *N* is slightly smaller than in Table 1.

summaries (see Table 1). For example, whereas the ratings of participants who completed the structured recall procedure (as a single intervention) yielded a medium effect size ($d = .55$) in the first measurement, a much smaller effect size was observed when the total valence of the retrieved items was considered ($d = .13$).

The effect size for the model-generating intervention was in the opposite direction when valence of the decision models was examined ($ds = .61$ and $-.16$ for the final ratings and for the valence, respectively). Furthermore, the Pearson correlation between the final ratings and the valence of the retrieved information was very low ($r = .08$), indicating that the ratings were unrelated to the retrieved information. The correlation between the final judgments and the valence of the decision models was somewhat higher ($r = .33$).

Prolonged Effect

As the interventions and the instructions to disregard did not serve to reduce reliance on preliminary information, examining their prolonged effect seems to be unwarranted. However, we carried out this measurement as part of the experimental design. There were no statistically significant differences between the first measurement ratings of participants who completed the experiment ($N = 59$; return rate of 35%) and participants who dropped out ($N = 109$), $t(238) = -0.08$, *ns*.

The effect sizes were similar to those obtained for the first measurement, suggesting that the prolonged effects of the interventions and the instructions to disregard do not vary much over time. The low return rate and the uneven division of participants between the experimental conditions make it difficult to reach any conclusions from this measurement.

Discussion

The present study focused on interventions for reducing the effect of preliminary information regarding a job applicant on evaluation of their subsequent performance. Specifically, we introduced information regarding a job applicant's abilities and examined whether it would influence how HR managers would rate his assessment-center summary. Such an influence might slant hiring decisions. HR managers' reliance on the preliminary information persisted, despite instructions to disregard it and interventions for reducing its effect.

Several conclusions can be drawn from the present work, mostly converging to demonstrate that preliminary information has an effect on subsequent

assessments. First, asking participants to disregard preliminary information—and even rewarding them for doing so—did not deter them from relying on the preliminary information in their final ratings of an applicant. Second, interventions for reducing reliance on the preliminary information were ineffective: Final ratings of the applicants' performance relied on the preliminary information to the same extent, regardless of whether or not the participants completed an intervention or were in the control group. Third, the interventions were executed in a satisfactory manner. The participants retrieved information and generated decision models that were virtually unmarred by the preliminary information. The collision between these two conclusions carries a potent organizational meaning: There was a dissociation between the interventions and the final ratings. This conclusion is supported by correlations between the final ratings and the valence of the retrieved information and the decision models, which were relatively low. Finally, contrary to predictions, participants' reliance on the preliminary information under the combined intervention was similar to that observed in the other conditions, which sheds light on expert performance.

This discussion opens with instructions to disregard and with the intricacies of expert judgments. We then examine reliance on preliminary information and its underlying mechanisms, and offer explanations for our most interesting finding: the dissociation between the exclusion of preliminary information from the execution of the interventions, and its presence in the final ratings. We conclude by highlighting some of the characteristics of expert judgment and making implications for personnel selection practices.

Instructions to Disregard

The futility of the instructions to disregard in the present study is not surprising, given a few characteristics of the preliminary information to be disregarded that habitually impair disregarding. First, the content of the preliminary information was closely related to the assessment-center summary that the participants read, rendering the implications of the preliminary information and the specific selection output difficult to separate (Wyer & Srull, 1986, 1989). Second, the preliminary information offers a causal structure for interpreting the assessment-center summary (Johnson & Seifert, 1994). Third, the motivation for disregarding the preliminary information was procedural. Individuals excluded evidence that was allegedly mistaken, but not evidence that was to be disregarded for reasons of confidentiality (Golding, Fowler, Long, & Latta, 1990; also see Sommers & Kassin, 2001). These conditions often prevail in the personnel selection scene, which renders relinquishment of preliminary information rather difficult.

This can be interpreted in the broader context of the tendency to rely on preliminary information.

Reliance on Preliminary Information

Reliance on preliminary information is a robust phenomenon (Chapman & Elstein, 2000; Jonas et al., 2001; Traut-Mattausch, Schulz-Hardt, Greitemeyer, & Frey, 2004) and it was found also in the present study. Such reliance may be beneficial from several perspectives: Adhering to directional preliminary information, which supposedly represents substantial knowledge, as did the preliminary information in our study, is a good Bayesian strategy (Koehler, 1991, 1993), as well as an adaptive one (Arkes, 1991; Friedrich, 1993). From an information-processing perspective, the race model (Feldman & Lynch, 1988) posits that memory does not necessarily dictate judgment. Rather, it can go the other way around, or the two processes might be altogether unrelated. This is relevant to the present findings because of the apparent preference to base judgments on general evaluations, such as the ones included in the preliminary information that we presented to participants.

It is plausible for an overall appraisal of a candidate and an assessment of his or her performance in a specific test to be positively correlated so that considering one while examining the other seems to be reasonable. However, it is doubtful that individuals can assign the preliminary information the “correct” weight, rather than rely on it overly, as indicated by ample studies (e.g., Ben-Shakhar et al., 1998). These considerations led us to try to reduce reliance on preliminary information by means of instructions to disregard, and an array of interventions.

Our instructions to disregard may have led inadvertently to cognitive effort via the process of *hyperaccessibility*: being constantly aware of information in order to ignore it deliberately (Wegner, 1992; Wegner & Erber, 1992). Likewise, thought suppression can have counterproductive effects, leading to arousal of the very thoughts or emotional states that the suppression is intended to prevent. Such backfiring is called an *ironic process* (e.g., Wenzlaff & Wegner, 2000).

Dissociation Between Interventions and the Final Ratings

The firm grip that preliminary information has on judgments calls for interventions to reduce this effect. However, such interventions are not always effective, and occasionally result in augmenting the effects that they

were intended to decrease. Research in accounting implies that mechanisms for enhancing the accuracy of audits (e.g., incentives, feedback, having to justify one's conclusions) can become stress inducers that have an adverse effect on performance (Ashton & Heiman, 1990). Likewise, instructions to ignore inadmissible evidence can have either curative or paradoxical effects on jury decisions (Kassin & Studebaker, 1998).

The interventions in the present study were designed to enable a comprehensive view of candidates' advantages and shortcomings by tackling the processes of information retrieval and generating a decision model. For the most part, the participants executed the interventions with hardly a trace of the preliminary information. Even when this information was present in the retrieved data, its presence was subdued, compared with the final ratings of the candidate. Thus, on the surface, it seems as though all of the prerequisites for impartial decision making were fulfilled. Nevertheless, the correlations between the final ratings and the valence of the interventions were low. This indicates that when individuals made the final ratings, they did not implement the techniques that were ingrained in the interventions that they had just executed. This dissociation between the interventions and the final judgments raises doubts as to whether the participants achieved any comprehensive insight.

The conundrum of this dissociation is not unprecedented. Previous research has shown that introducing explicit requirements for disregarding, in settings in which the presence or absence of the preliminary information can be detected immediately, leads to a fulfillment of these requirements at the cost of later disobeying them altogether. This rebound effect was demonstrated when repressed stereotypes emerged in memory a week after the initial repression task, taking the place of nonstereotypical items (Macrae, Bodenhausen, Milne, & Wheeler, 1996). Our interventions may have caused such an adverse effect.

In a closer domain, note taking during the selection interview improved detailed recall of interview data, but did not reduce the effect of pre-interview impressions on the final ratings (Macan & Dipboye, 1994). A direct link between the decision aid and the decision helped: Consulting the notes while performing the judgment led to an increase in judgment accuracy (Middendorf & Macan, 2002).

Slowiaczek, Klayman, Sherman, and Skov (1992) concluded, "Efficient information gathering does not guarantee efficient or unbiased use of the information gathered" (p. 392). This implies that interventions may lead to an accessibility of confirmatory and contradictory information without making a dent in the judgment process, perhaps because participants perceived the information gathering and the judgment as separate tasks. Perhaps there must be a distinct linkage between the information-gathering phase

(which in the context of the present study also includes model generating) and the judgment.

Expert Decision Processes

We hypothesized that the combined intervention condition (retrieving positive and negative information about the applicant and then generating a model for determining his rating) would be associated with the most intensive reliance on preliminary information as a result of cognitive overload (Higgins & Liberman, 1994). Previous research has demonstrated the aversive effect of cognitive overload on the use and unconscious choice of various mental tools (Fiske & Taylor, 1991; Seale & Rapoport, 2000). Specifically, according to Stangor and McMillan (1992), "congruency effects are most likely to occur when processing occurs under cognitively demanding conditions" (p. 58).

Yet, our hypothesis was not confirmed. Participants' reliance on the preliminary information under this intervention did not exceed their reliance under other interventions, presumably because our participants were experts who have elaborate cognitive processes available. These processes include knowledge, availability of schemes and scripts, ability to quickly process information, and a capacity to make versatile use of all of these qualities according to circumstances (Ganzach, 1994; Showers & Cantor, 1985). Knowledge and understanding enable experts to sift through a multitude of stimuli, segregating them into relatively large meaningful pieces of information (Chase & Simon, 1973). They also allow experts to recognize patterns that refer to goal-relevant constraints (Vicente & Wang, 1998). Ericsson and Lehmann's (1996) review indicated that through practice and feedback, experts acquire vast knowledge, which is organized in patterns and in mental models that facilitate quick retrieval and problem solving.

Having given the experts such extensive credit, one cannot help but wonder why they did not use their superior skills in a manner that would eliminate the preliminary information from their final ratings. Yates, Veinott, and Patalano (2003) shed light on this issue, suggesting that decision aids relate to a narrow scope of decision quality; whereas for decision makers, this concept is much broader. Thus, decision aids are often ignored, as they seem to be irrelevant to significant decider concerns. Further supporting this claim, the literature on expert decision making indicates that experts are often rewarded for properties other than decision accuracy; for example, appearing to possess singular knowledge (Shanteau, 1988). The literature also indicates that experts' judgments are reached by more elaborate processes than those of laypeople, yet they do not necessarily have better predictive value or, as in

our case, a better ability of excluding the preliminary information from the final ratings (Camerer & Johnson, 1991).

Experience in personnel selection was not associated with our expert participants' adherence to instructions to disregard. This is perhaps contrary to common intuition equating experience with enhanced expertise and improved judgment, yet very much in line with previous findings attesting to the lack of a relation between these variables (e.g., Ashton, 1974, 2000; Einhorn & Hogarth, 1978). Fields such as weather forecasting (Winkler & Murphy, 1973) or horse-race commentating (Dowie & Elstein, 1988, as cited in Fischhoff, 1982) provide experts with frequent and calibrating feedback. On the other hand, HR managers seldom receive immediate and unequivocal feedback on their decisions, either because the false negatives are not recruited or because of the lack of clear criteria for personnel evaluation.

Implications for Procedures Promoting the Relinquishment of Preliminary Information

The present findings, as well as a large body of research, point to the influential role that preliminary information plays in lay and expert decision making across a plethora of domains. Reducing the tendency to rely on preliminary information in personnel selection goes beyond a theoretical issue, as hiring procedures are designed to enable a systematic interpretation of professional input, with every phase preferably not being contingent on the previous one.

Several conditions could facilitate disregarding preliminary information, out of which two appear to be most pertinent to our paradigm. First, Wyer and Srull (1986, 1989) proposed that information is stored in bins, according to topic, and is retrieved by cues pertaining to the header of that bin. A retrieval cue less related to the preliminary information, such as what the applicant's strengths would be as an employee (rather than considering him as an applicant, consistent with the framing of preliminary information), might prevent reliance on preliminary information during the final rating.

Second, excessive instructions to disregard may have had a boomerang effect. Telling mock jurors that evidence was inadmissible led to it being disregarded, whereas an additional admonishment from the judge enhanced both the influence of the evidence on the verdicts and the participant's desire to consider the evidence (Wolf & Montgomery, 1977; also see Macrae et al., 1995, 1996). The underlying mechanism could be the hyperaccessibility discussed previously. It would be beneficial to examine how participants incorporate (or disregard) preliminary information in their judgments when they are told only once to disregard it, rather than being constantly warned against regarding it.

Finally, exposure to biasing factors, which, as we have demonstrated, can be relevant and seemingly harmless, has been described as contaminating (Wilson, Centerbar, & Brekke, 2002). As successful decontamination is extremely hard to achieve, it might be more realistic to aim at procedures in which decision makers are only exposed to the segment of information that they are asked to assess.

The present study demonstrated the tenacious hold that preliminary information has on judgments, despite various mechanisms for reducing this effect and despite the fact that the participants were experienced HR managers, who presumably are capable of putting preliminary information aside. The importance of our findings lies in avoiding exposure to preliminary information and in emphasizing the importance of reducing its effect in the event that one is exposed to it.

References

- Arkes, H. R. (1991). Costs and benefits of judgmental errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486–498.
- Ashton, R. H. (1974). Cue utilization and expert judgments: A comparison of independent auditors with other judges. *Journal of Applied Psychology*, *59*, 437–444.
- Ashton, R. (2000). A review and analysis of research on the test–retest reliability of professional judgment. *Journal of Behavioral Decision Making*, *13*, 277–294.
- Ashton, R. H., & Heiman, V. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback. *Journal of Accounting Research*, *28*, 148–180.
- Baltes, B. B., & Parker, C. P. (2000). Reducing the effects of performance expectations on behavioral ratings. *Organizational Behavior and Human Decision Processes*, *82*, 237–267.
- Ben-Shakhar, G., Bar-Hillel, M., Bilu, Y., & Shefler, G. (1998). “Seek and ye shall find”: Test results are what you hypothesize they are. *Journal of Behavioral Decision Making*, *11*, 235–249.
- Burnstein, E., & Schul, Y. (1982). The informational basis of social judgments: Operations in forming an impression of another person. *Journal of Experimental Social Psychology*, *18*, 217–234.
- Camerer, C. F., & Johnson, E. J. (1991). The process–performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 195–217). Cambridge, UK: Cambridge University Press.

- Chapman, G. B., & Elstein, A. S. (2000). Cognitive biases in medical decision making. In G. B. Chapman & F. A. Sonnenberg (Eds.), *Decision making in health care: Theory, psychology, and applications. Cambridge series on judgment and decision making* (pp. 183–210). New York: Cambridge University Press.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous diagnostic observations. *Journal of Abnormal Psychology, 72*, 193–204.
- Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74*, 271–280.
- Chapman, L. J., & Chapman, J. P. (1982). Test results are what you think they are. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 239–248). New York: Cambridge University Press.
- Chase, W. G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.
- Cloyd, C. B., & Spilker, B.C. (1999). The influence of client preferences on tax professionals' search for judicial precedents, subsequent judgments, and recommendations. *Accounting Review, 74*, 299–322.
- Cloyd, C. B., & Spilker, B.C. (2000). Confirmation bias in tax information search: A comparison of law students and accounting students. *Journal of American Taxation Association, 22*(2), 60–71.
- Cohen J. E. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582.
- Dowie, J. & Elstein, A.S. (Eds.) Professional judgment: A reader in clinical decision making. New York: Cambridge University Press.
- Einhorn, H. J., & Hogarth, R .M. (1978). Confidence in judgment: Persistence in the illusion of validity. *Psychological Review, 85*, 395–416.
- Ericsson, A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptations to task constraints. *Annual Review of Psychology, 47*, 273–305.
- Evans, J. St. B. T., Newstead, S. E., Allen, J. L., & Pollard, P. (1994). Debiasing by instruction: The case of belief bias. *European Journal of Cognitive Psychology, 6*, 263–285.
- Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology, 73*, 421–435.
- Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York: Cambridge University Press.

- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw-Hill.
- Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, *100*, 289–319.
- Ganzach, Y. (1994). Theory and configularity in expert and layperson judgment. *Journal of Applied Psychology*, *79*, 439–448.
- Golding, J. M., Fowler, S. B., Long, D. L., & Latta, H. (1990). Instructions to disregard potentially useful information: The effects of pragmatics on evaluative judgments and recall. *Journal of Memory and Language*, *29*, 212–227.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258–268.
- Higgins, E. T., & Liberman, A. (1994). Memory errors from a change of standard: A lack of awareness or of understanding? *Cognitive Psychology*, *27*, 227–258.
- Hirt, E. R., & Markman, K. D. (1995). Multiple explanations: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology*, *69*, 1069–1086.
- Johnson, H. M., & Seifert, C. M. (1994). Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 1420–1436.
- Jonas, E., Frey, D., & Greenberg, J. (2003). *Decision vs. information focus in information seeking: Factors that affect the preference for supporting and conflicting information after decisions*. Working paper, University of Munich.
- Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology*, *80*, 557–571.
- Kassin, S. M., & Studebaker, C. A. (1998). Instructions to disregard and the jury: Curative and paradoxical effects. In C. M. MacLeod & J. M. Golding (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 413–434). Mahwah, NJ: Lawrence Erlbaum.
- Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, *110*, 499–519.
- Koehler, J. J. (1993). The influence of prior beliefs on scientific judgments of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*, 28–55.

- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 107–118.
- Lewis, A., & Sherman, S. J. (2003). Hiring you makes me look bad: Social identity-based reversals of the in-group favoritism effect. *Organizational Behavior and Human Decision Processes*, *90*, 262–276.
- Macan, T. H., & Dipboye, R. L. (1990). The relationship of interviewers' pre-interview impressions to selection and recruitment outcomes. *Personnel Psychology*, *43*, 745–768.
- Macan, T. H., & Dipboye, R. L. (1994). The effects of the application on processing of information from the employment interview. *Journal of Applied Social Psychology*, *24*, 1291–1314.
- Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*, 397–407.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Wheeler, V. (1996). On resisting the temptation for simplification: Counterintentional effects of stereotype suppression on social memory. *Social Cognition*, *14*, 1–20.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the literature*. Minneapolis, MN: University of Minnesota Press.
- Middendorf, C. H., & Macan, T. H. (2002). Note-taking in the employment interview: Effects on recall and judgments. *Journal of Applied Psychology*, *87*, 293–303.
- Moser, K., Gadenne, V., & Schröder, J. (1988). Under what conditions does confirmation seeking obstruct scientific progress? *Psychological Review*, *95*, 572–574.
- Murthi, B. P. S., & Srinivasan, K. (1999). Consumers' extent of evaluation in brand choice. *Journal of Business*, *72*, 229–256.
- Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, *29*, 261–265.
- Philips, A. P., & Dipboye, R. L. (1989). Correlational tests of predictions from a process model of the interview. *Journal of Applied Psychology*, *74*, 41–52.
- Saad, S., & Sackett, P. R. (2002). Investigating differential prediction by gender in employment-oriented personality measures. *Journal of Applied Psychology*, *87*, 667–674.
- Sawyer, J. (1966). Measurement and prediction: Clinical and statistical. *Psychological Bulletin*, *66*, 178–200.
- Schulz-Hardt, S., Frey, D., Luthgens, C., & Moscovici, S. (2000). Biased information search in group decision making. *Journal of Personality and Social Psychology*, *78*, 655–669.

- Seale, D. S., & Rapoport, A. (2000). Optimal stopping behavior with relative ranks: The secretary problem with unknown population size. *Journal of Behavioral Decision Making, 13*, 391–411.
- Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. *Acta Psychologica, 68*, 203–215.
- Showers, C., & Cantor, N. (1985). Social cognition: A look at motivated strategies. *Annual Review of Psychology, 36*, 275–305.
- Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research, 7*, 99–111.
- Slowiaczek, L. M., Klayman, J., Sherman, S. J., & Skov, R. B. (1992). Information selection and use in hypothesis testing: What is a good question, and what is a good answer? *Memory and Cognition, 20*, 392–405.
- Sommers, S. R., & Kassin, S. M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin, 27*, 1368–1377.
- Stangor, C., & McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: A review of the social development literatures. *Psychological Bulletin, 111*, 42–61.
- Terpstra, D. E., & Kethley, R. B. (2002). Organizations' relative degree of exposure to selection discrimination litigation. *Public Personnel Management, 31*, 277–292.
- Traut-Mattausch, E., Schulz-Hardt, S., Greitemeyer, T., & Frey, D. (2004). Expectancy confirmation in spite of disconfirming evidence: The case of price increases due to the introduction of the Euro. *European Journal of Social Psychology, 34*, 739–760.
- Vicente, K. J., & Wang, J. H. (1998). An ecological theory of expertise effects in memory recall. *Psychological Review, 105*, 33–57.
- Wegner, D. M. (1992). You can't always think what you want: Problems in the suppression of unwanted thoughts. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 193–225). San Diego, CA: Academic Press.
- Wegner, D. M., & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology, 63*, 903–912.
- Wenzlaff, R. M., & Wegner, D. M. (2000). Thought suppression. *Annual Review of Psychology, 51*, 59–91.
- Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185–200). New York: Cambridge University Press.
- Winkler, R. L., & Murphy, A. H. (1973). Experiments in the laboratory and in the real world. *Organizational Behavior and Human Performance, 10*, 252–270.

- Woehr, D. J., & Feldman, J. (1993). Processing objective and question order effects on the causal relation between memory and judgment in performance appraisal: The tip of the iceberg. *Journal of Applied Psychology, 78*, 232–241.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology, 67*, 189–205.
- Wolf, S., & Montgomery, D. A. (1977). Effects of inadmissible evidence and level of judicial admonishment to disregard on the judgments of mock jurors. *Journal of Applied Social Psychology, 7*, 205–219.
- Wyer, R. S., Jr., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review, 93*, 322–359.
- Wyer, R. S., Jr., & Srull, T. K. (1989). *Memory and cognition in its social context*. Hillsdale, NJ: Lawrence Erlbaum.
- Yates, F., Veinott, E. S., & Patalano, A. L. (2003). Hard decisions, bad decisions: On decision quality and decision aiding. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 1–63). New York: Cambridge University Press.