

On Wining: The Impact of Scarcity on Allocators' Evaluations of Candidates for a Resource

MICHAEL ROSS AND JOHN H. ELLARD

University of Waterloo and University of Michigan

Received November 20, 1985

It was hypothesized that allocators unwittingly forge a relation between the availability of a resource and the worthiness of the applicants for that resource. In a simulation of an academic job search, graduate students were instructed to hire either 2 or 6 of 12 excellent candidates. Subjects spoke into a tape recorder reporting their initial reactions to each of the candidate's vitae, then made their hiring decisions, and rated the desirability of each of the candidates. Next, they were told they could now hire more of the applicants if they wished to do so. In a control condition, subjects were treated identically to experimental subjects in all aspects of the procedure but one—they did not anticipate and were not required to make hiring decisions. Instead, they rank ordered the applicants in terms of quality with the knowledge that "their department" wished to hire either 2 or 6 people. As hypothesized, subjects anticipating 2 appointments made fewer positive comments about the candidates on the audiotapes than did those anticipating 6 appointments; subjects in the experimental condition evaluated the candidates they hired more favorably than subjects in the control condition evaluated their own top-ranked candidates; subjects who were initially required to hire 2 candidates continued to employ fewer applicants than those who initially hired 6 when external hiring constraints were removed. The implications of the data for the review process in academic psychology journals were discussed.

© 1986 Academic Press, Inc.

When a resource is scarce, society must determine how much of it will be made available, within the constraints of natural scarcity and conflicting priorities, and who should receive it (Calabresi & Bobbitt, 1978). We refer to this latter decision as the winnowing process. The purpose of the present research is to examine the facility with which

This research was supported by a Social Sciences and Humanities Research Council of Canada (SSHRC) Research Grant and Leave Fellowship to Michael Ross and a SSHRC Postdoctoral Fellowship to John Ellard. We thank John Holmes, Mel Lerner, Melvin Manis, Jim Olson, Gerry Sande, Yaacov Schul, Kris Robinson Staveley, Erik Woody, and Mark Zanna for their comments on an earlier version of the manuscript. Requests for reprints should be sent to Michael Ross, Psychology Department, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

decision makers accept the results of winnowing and the implications of this acceptance.

Our basic premise is that various psychological processes induce allocators to create a dependence between the availability of a resource and the perceived worthiness of the applicants for the resource. Some examples may help elucidate this principle. The number of academic jobs available in a psychology department typically depends more upon departmental needs and economic factors than on the quality or availability of applicants. Similarly, grant review panels have relatively meager amounts of money to allocate, and journal editors work within page limitations. It is probably fair to say that these limitations primarily reflect economic and practical concerns, rather than a rigorous analysis of the number of "meritorious" applicants (see, e.g., Beyer (1978) for a discussion of the practices of scientific journals).

Nonetheless, we propose that decision makers may come to view their allocation of the resource as conditioned primarily or even solely by the worthiness of the applicants. For instance, allocators may assume that these job applicants, grant requests, and manuscripts would be acceptable and those objectionable, almost regardless of monetary constraints, page limitations, and so forth. In short, judgments of relative worthiness may be converted to judgments of absolute worthiness, with the result that the general availability of the resource is seen as a reflection of the need or worthiness of the applicants.

In real life, decision makers may avail themselves of a variety of winnowing procedures. These include lotteries and allocations based upon ascribed characteristics of the recipients, such as race, gender, and pedigree. In this paper, we limit our analysis to equal opportunity allocations in which decision makers attempt to distribute the resource on the basis of the relative worthiness of the applicants. Our analysis suggests that, in these latter allocations, decision makers tend to view their distribution of the resource as fair and just. We first consider the bases of this judgment, and then its implications.

The Perception of Fairness

Decisions regarding the general availability and distribution of a scarce resource are often effected by different people. For example, personnel officers are assigned a set of positions to fill. In such instances, resource availability is not the direct concern of the allocators. Their focus is on the winnowing process. In essence, they adopt a problem solving orientation. They employ procedures and criteria that will permit them to allocate the resource.

Moreover, allocators may typically view their decisions as justifiable. There are several possible determinants of this perception that the winnowing process yields just outcomes. First, allocators are educated to

accept the suitability of particular allotment procedures (Lerner, 1980). Second, allocators can often influence the selection of winnowing criteria and/or their application. This degree of freedom probably increases the perceived legitimacy of the decisions (Thibaut & Walker, 1975).

In the present research we focus on two additional factors that may influence decision makers' approval of the winnowing process. It is hypothesized that the general availability of the resource may affect the standards and evaluation strategies employed by the allocators. The severity of standards may vary inversely with availability. When a university faces a paucity of new faculty appointments, the hiring and tenure decisions are likely to be more stringent than when positions are readily available. More subtly, resource availability may alter the strategies allocators utilize and the information they seek in their assessments of worthiness. When the resource is relatively abundant, allocators may search for positive features, merits that would entitle an applicant to a resource, as well as faults. In contrast, when the resource is extremely scarce, allocators need to exclude the vast majority of applicants. This requirement may induce a focus on flaws. Evidence of nonworthiness will warrant rejection of an applicant. Note that it may not be necessary for allocators to identify a host of flaws in scarce resource conditions. Conceivably, a relatively few faults can serve as a basis of rejection when a resource is extremely limited.

This analysis of the impact of resource availability on standards and evaluation strategies has an important implication. Standards and strategies may shift so as to provide allocators with support for their decisions. In scarce resource conditions, the applicants' flaws are salient; it may thus seem that most of them do not deserve the resource, anyway. In more abundant resource conditions, the applicants' merits also stand out; consequently, a more generous distribution of the resource may seem fair.

Finally, once a decision has been made, concerns for consistency and justice may prompt cognitive changes. Allocators can justify their actions, postdecisionally, by making their attitudes toward the applicants fully consonant with their choices (Aronson, 1969; Festinger, 1957; Heider, 1958; Olson & Zanna, 1982; Wicklund & Brehm, 1976). Successful applicants may come to be seen as highly deserving of the resource and unsuccessful applicants as unworthy of it (Lerner, 1980).

Absolute vs. Relative Worthiness

It seems likely that allocators are generally unaware of the impact of the availability of the resource on their standards, evaluation strategies, and postdecisional attitudes. If so, the above analysis implies that judgments of relative worthiness may be readily converted to assessments of absolute worthiness. It is our thesis that allocators tend to link their assessment of the fairness of the distribution process to their evaluation of the

fairness of the outcome. As a consequence, they perceive the resource as sufficiently abundant that most who deserve to obtain it, do so. In general, then, they tend to underestimate the impact of the overall availability of the resource on their decisions.

A failure to maintain a distinction between absolute and relative judgments of worthiness has at least two insidious implications. First, allocators may fail to agitate for an increase in total resources because the current allotment procedure is seen as entirely defensible. When the resource pool is unfairly low, however, it is important that allocators do the agitating. The only other probable source of discontent is the losers of the competition and their demands are typically rejected as self-serving.

In addition, consider the attributional significance of underestimating the impact of external constraints on allocations. For example, some of the major journals in psychology reject over 80% of the manuscripts submitted. Such rejection rates reflect, in part, severe page limitations and the desire of editors to keep publication lags tolerable (Beyer, 1978). If editors and reviewers do not attribute the rejection rate, at least in part, to a shortage of space, they may attribute it either to the quality of the research, or to the nature of psychology, itself. For instance, one might conclude that most research being conducted today is seriously flawed; alternatively, one might assume that psychology is a highly subjective, nonscientific discipline in which intelligent people disagree on what is worth doing. Yet neither of these judgments may be entirely valid. If rejection rates could be reduced through the easing of page restrictions, then perceptions of the quality and subjectivity of the enterprise might be altered.

The Current Study

The major purpose of the present research was to examine our thesis that (a) the general availability of the resource affects allocators' evaluation strategies, and (b) allocators' judgments of relative worthiness are converted to evaluations of absolute worthiness. Our experimental test involved a simulation of an academic job search. Graduate students were shown the vitae of 12 excellent applicants. The students played the role of members of a hiring committee of a rapidly expanding department of psychology. In the experimental conditions, subjects were instructed either to hire two or six of the candidates. Subjects spoke into a tape recorder while initially reading the candidates' vitae, reporting their reactions to each. After perusing the vitae and making their hiring decisions, subjects rated the overall desirability of each of the candidates on bipolar scales. Next, experimental subjects were informed that their department had received additional funds. They could now hire as many more of the applicants as they wished. They were told, however, that they should hire only good people, individuals they would like to see in their own

department. This final hiring session provided a behavioral measure of subjects' tendency to convert evaluations of relative worthiness to assessments of absolute worthiness.

The experiment also included control conditions. Control subjects were treated identically to experimental subjects in all aspects of the procedure, but one. In contrast to experimental subjects they did not anticipate and were not required to make hiring decisions.

From our assessment of the dynamics of the winnowing process, we derived the following hypotheses and predictions:

(1) Resource availability affects evaluation strategies. Our analysis suggests that coding of the audiotapes should reveal that experimental and control subjects express fewer positive comments about the applicants when two positions are available than when six are available. The impact of number of positions available on frequency of negative comments is less evident. As noted above, when resources are scarce, a few faults may be sufficient to establish evidence of nonworthiness. A stronger prediction can, thus, be made with respect to positive than with respect to negative comments.

(2) On the face of it, decisions seem more difficult as scarcity increases. Our analysis suggests, however, that this may not always be the case. If flaws can be seen as a basis for rejecting a candidate, then decisions can be made with some ease when resources are very limited. Although many candidates may be excellent, few are perfect. Decisions may actually be more complex when the resource is somewhat less scarce and allocation judgments require a careful weighting of the pluses and minuses of each candidate. Hence our prediction is that subjects will view their initial hiring decisions as more difficult when six positions are available than when two are available—even when the number of excellent candidates clearly exceeds six.

(3) Postdecisional consistency pressures induce allocators to bring their attitudes in line with their decisions. There are two possible foci for a postdecisional shift in attitudes. Experimental subjects could exaggerate the merits of candidates that they hire and/or derogate the applicants that they reject. There is reason to believe that the enhancement effect may be stronger than the derogation effect. As Gerard and White (1983) noted, "It is the chosen alternative that the person has to live with and accommodating to it becomes a paramount concern. The rejected alternative is a foregone option" (p. 367). We assessed a postdecisional shift in attitudes by contrasting the ratings experimental subjects provided for those they hired and rejected to the ratings of comparably ranked applicants by control subjects who were not required to make a hiring decision.

(4) Finally, we expected the combination of changes in evaluation strategies and postdecisional shifts in attitude to facilitate conversion of judgments of relative merit to assessments of absolute merit. This implies

that allocators will tend to stand by their initial evaluations when the resource becomes plentiful. Thus, when subjects are informed that they can now employ as many additional applicants as they wish, those who could originally hire only two should end up employing fewer candidates than those who could originally hire six.

METHOD

Subjects and Design

The subjects were 25 female and 21 male psychology graduate students recruited by a fellow student. Only one of the students approached refused to be in the study. Subjects were paid \$10.00 for their contribution. They participated individually in one session of 2 hr and were randomly assigned to the four conditions.

Stimulus Materials

The 12 curriculum vitae used for the study were selected from those submitted by approximately 60 applicants for an actual job opening in social psychology at the University of Waterloo in the fall of 1981. Although there was variability in quality among the selected vitae, all 12 candidates were first-rate in our view. Careful attention was given to maintaining the confidentiality of the actual applicants through extensive changes to each of the vitae. The names, publications, and relevant experience of applicants were altered. The 12 candidates were assigned male names to hold gender constant as a factor in subjects' evaluations.

Procedure

Subjects were informed that they were participating in a study of the decision making process. They were told that they would be reviewing and evaluating the curriculum vitae of 12 applicants for positions in a psychology department. They were asked to role-play members of a hiring committee. Preliminary instructions emphasized that subjects would be under time pressure to complete all aspects of the procedure in the 2-hr allotted time period. To assist subjects in managing their time, the experimenter suggested that they not spend more than 3 or 4 min, on average, reviewing each of the vitae.¹ The experimenter also indicated that subjects were free to make brief notes as they evaluated the candidates.

Subjects were then left alone to follow written step-by-step instructions for the balance of the session. After reading a brief overview of the procedure, all subjects read a detailed description of the department in which the competition was being held. The successful candidates in the competition would be the first social psychologists in a new and rapidly growing psychology department; they would be expected to develop and maintain a strong program in social psychology. The department was interested in hiring well-rounded scholars with demonstrated research, teaching, and administrative potential. Since the department was still growing, special talents or training in areas outside of social psychology such as statistics or survey methodology would be looked upon favorably. The department was stated to be strong in developmental and clinical psychology, with most of the current faculty members being at the assistant or associate level.

¹ Instructions encouraging subjects to manage their time efficiently were found to be necessary during pilot testing of the procedure. The procedure initially called for subjects to evaluate 20 vitae in the same period. Subjects' intrinsic interest in the stimulus materials and concern for carrying out a competent evaluation resulted in a high degree of frustration and a 3- to 4-hr session. With the number of vitae cut back to 12, subjects reported that they found the procedure taxing but manageable; most completed the task within 2 hr.

All subjects then read detailed instructions for the evaluation phase. One set of instructions described the think aloud procedure. Subjects were instructed to start an audio cassette tape recorder as soon as they began evaluating the vitae and to "verbalize ongoing thoughts as they occur." Subjects were encouraged to record their personal reactions to each of the vitae. All subjects received the 12 vitae in the same order.

A second set of instructions directed subjects in the experimental conditions to hire either 2 or 6 of the applicants after reading all 12 vitae. Experimental subjects were provided with a "Hiring Decision" form that listed the names of the candidates. Subjects were asked to place a check beside each name in either the "Hire" or "Not Hire" column. Subjects read all of the vitae before making the hiring decisions.

Experimental subjects then rated the overall quality of each candidate by placing a slash through 150-mm scales with end points labeled highly undesirable (1) and highly desirable (150). After evaluating the desirability of each candidate, experimental subjects were asked to reconsider the 12 candidates for positions in the same department. The instructions were as follows:

It is occasionally the case that applicants for a particular job competition are reviewed a second time as a result of new circumstances. New hiring funds may become available or successful candidates may choose to go elsewhere at the last minute. I would now like you to go through the hiring decision process once again assuming that such hypothetical circumstances had arisen. In this instance, I would like you to consider the hypothetical extreme: assume that this large department had *unlimited resources* and could hire *as many* candidates as it wanted to . . . Although the department now has unlimited resources, it is nevertheless committed to maintaining a high standard, so you should only hire candidates you consider to be *highly* qualified. At the extreme you could hire all 12 if you consider them all to be highly qualified. But make your decisions carefully. Hire only those you think are extremely promising and whom you would like to see as members of the department.

These decisions were recorded on a form that was identical to that used for the initial hiring. Finally, subjects evaluated the difficulty of their hiring decisions by putting a slash through a 150-mm line with end points labeled not at all difficult and extremely difficult. Separate scales were provided for the first and second hiring decisions.

Control subjects were asked to review the 12 candidates with the objective of rating the overall desirability of each, and with the knowledge that their department ultimately wished to hire either 2 or 6 of the applicants. After tape recording their initial reactions to the 12 vitae, control subjects rated each candidate on the same 150-mm scales completed by experimental subjects. Control subjects were then asked to rank order the 12 candidates, with no ties allowed. Their rankings of the candidates were used to create target groups of comparable quality and size to experimental subjects' "hired" and "rejected" groups.

Subjects were debriefed in 2- or 3- person groups by one of the authors. Subjects were encouraged to report their perceptions of the task, including their own reports of the strategies they used in evaluating the candidates. Their comments revealed no awareness of the hypotheses being tested.

RESULTS

Initial Reactions

All subjects tape recorded their initial reactions to the candidates' vitae. A research assistant, blind to experimental condition, listened to the tapes and tabulated the number of positive and negative comments

subjects recorded for each of the 12 candidates. A second assistant coded one-third of the tapes. The correlation between the ratings of the two assistants was .84.

We calculated the total number of positive and negative comments each subject made for the 12 candidates (see Table 1 for means). These totals were submitted to a 2 (experimental vs. control) \times 2 (2 appointments vs. 6) \times 2 (positive vs. negative comment) analysis of variance.² The first two factors were between-subject variables. The final factor was a within-subject variable. The analysis revealed that subjects expressed more positive comments, $M = 60.5$, than negative, $M = 35.8$, $F(1, 38) = 88.9$, $p < .001$. This presumably reflects, in part, the general excellence of the vitae. More interestingly, the interaction between number of appointments and the nature of the comment was significant, $F(1, 38) = 6.50$, $p < .025$. Subjects anticipating 2 appointments made fewer positive comments, $M = 55.4$, than did those anticipating 6 appointments, $M = 66.1$, $t(38) = 1.86$, $p < .05$ (one-tailed). On the other hand, the frequency of negative comments did not differ significantly as a function of the number of appointments, $t < 1$.

Note that comparable results were obtained in the control and experimental conditions. There were no significant main effects or interactions involving this variable. Thus, subjects' evaluation strategies were influenced by the number of candidates to be employed but unaffected by whether the subjects, themselves, would do the hiring.

The following comments provide examples of subjects' responses to the vitae:

This guy looks good. I don't know why I don't like him. He has got the best qualifications, I think, so far as publications and abilities. Maybe it is just the 'junior' after his name. I always think they are being pretentious when they put a junior after their name.

I like this "I'm prepared to teach" instead of just "I'm interested in". . . .

He says "I am prepared to teach". I'd rather see if he's interested to teach. Ok if you are prepared. I want more interest.

So this guy's got nothing out. Two in press, one in a good journal, one lousy . . . not very impressive.

Oh shit, one publication but he's got better teaching experience than guy number one.

² Four subjects were excluded from this analysis, one from each condition. The data were not available for three subjects because of tape recorder malfunctions; the fourth failed to follow the thought listing instructions.

TABLE 1
MEAN NUMBER OF POSITIVE AND NEGATIVE REACTIONS TO CANDIDATES' VITAE

| | Experimental | | Control | |
|------------------------|--------------|------|---------|------|
| | 2 | 6 | 2 | 6 |
| Number of appointments | | | | |
| Positive comments | 55.4 | 63.8 | 55.4 | 68.9 |
| Negative comments | 38.4 | 34.9 | 36.7 | 34.0 |

As these comments indicate, subjects reported candid and rather idiosyncratic reactions to the vitae. Some valued applied or teaching experience; others did not. Some counted number of publications; others emphasized quality over quantity and so forth. Clearly there was plenty of room for personal criteria to play a role in the evaluation process.

Quality Ratings

After making their hiring decisions, subjects in the experimental conditions evaluated the desirability of each of the applicants. Control subjects made similar ratings prior to ranking the candidates. We compared the evaluations control subjects reported for their top 2 (2-appointment condition) or top 6 (6-appointment condition) ranked candidates to the evaluations experimental subjects assigned to the candidates they hired. This procedure compares experimental and control subjects' evaluations of the candidates they judged to be superior. A 2×2 between-subjects analysis of variance revealed two significant main effects. Subjects in the experimental conditions made more favorable evaluations, $M = 123.23$, than subjects in the control conditions, $M = 112.94$, $F(1, 42) = 5.19$, $p < .05$. In addition, subjects' evaluations of the top 2 candidates in the 2-appointment conditions exceeded the evaluations of the top 6 candidates in the 6-appointment conditions, $F(1, 42) = 11.04$, $p < .01$, M s of 125.71 and 110.70, respectively (which is not surprising, as the 6-appointment condition includes less desirable candidates).

Comparable analyses were conducted on ratings of candidates who were either not hired by the experimental subjects or who were ranked by control subjects lower than 2 (2-appointment condition) or 6 (6-appointment condition). There were no significant effects in mean evaluations of these inferior candidates. Most important, the difference in the quality ratings of experimental, $M = 70.5$, and control, $M = 68.0$, subjects did not approach significance, $F < 1$.

Second Hiring Decision

We examined the second hiring decision to assess the number of candidates employed in the experimental conditions when external hiring

TABLE 2
PERCEIVED DIFFICULTY OF HIRING DECISIONS

| | Number of appointments | |
|------------------------|------------------------|------|
| | 2 | 6 |
| First hiring decision | 65.6 | 83.2 |
| Second hiring decision | 94.6 | 85.1 |

Notes. (1) The data are from experimental subjects only. (2) Scale values range from 1 to 150, with larger values indicating more difficulty.

constraints were removed. As anticipated, subjects in the 2-appointment condition would have hired fewer applicants under conditions of abundance, $M = 6.23$, than subjects in the 6-appointment condition, $M = 8.08$, $t(18.6) = 3.46$, $p < .01$.³

Perceived Difficulty of the Hiring Decisions

At the conclusion of the study, experimental subjects rated the difficulty of their hiring decisions on two bipolar scales. The mean ratings are shown in Table 2. A 2×2 ANOVA was conducted on these data. Experimental condition (2 appointments vs. 6) was a between-subjects variable and hiring decision (first vs. second) a within-subjects variable.

The ANOVA yielded a main effect for hiring decision, $F(1, 23) = 5.33$, $p < .05$. Subjects rated the second hiring decision as more difficult than the first. This main effect is qualified by a marginally significant interaction, $F(1, 23) = 4.09$, $p = .055$. As predicted, subjects in the 6-appointment condition judged the first hiring decision to be more difficult than did subjects in the 2-appointment condition (see Table 2). In contrast, subjects in the 2-appointment condition reported greater difficulty in making the second hiring decisions than did subjects in the 6-applicant condition. This latter trend may simply reflect the fact that subjects in the 2-appointment condition had a greater number of candidates to reconsider during the second hiring decision. It should be noted, however, that simple effects analysis revealed that neither of these trends attained statistical significance.

Regression Analyses

We examined whether subjects' preliminary reactions to the candidates (obtained from the audiotapes) predicted their initial hiring decisions,

³ The degrees of freedom used were fractional because the variances were heterogeneous and a separate variance t value was calculated. The procedure is described in Myers (1979).

their quality ratings, and second hiring decisions. Multiple regression analyses were performed separately in the experimental and control conditions. In all of the analyses, the frequency of positive and negative comments were entered as predictors. In the experimental condition, the multiple R was significant for the initial hiring measure, $F(2, 20) = 4.57, p < .025 (R^2 = .31)$, the quality ratings, $F(2, 20) = 5.16, p < .025 (R^2 = .34)$, and marginally significant for the second hiring measure, $F(2, 20) = 2.60, p < .10 (R^2 = .21)$. In the control condition, the multiple R relating initial reactions to quality ratings was significant, $F(2, 16) = 4.98, p < .025 (R^2 = .38)$.

The standardized regression coefficients, β , obtained from the regression analyses are shown in Table 3. As Table 3 reveals, the frequency of positive comments was a significant predictor of all of the criteria; whereas, the frequency of negative comments was a weaker predictor in the experimental condition. It is important to emphasize, though, that direct contrasts indicated that positive and negative comments did not differ significantly in terms of the absolute strength of their relationship to the various criteria (ignoring sign).

A second set of analyses was conducted entering number of candidates to be hired (2 vs. 6) as a dummy variable. The results reported above were not altered significantly by this addition.

In summary, then, the regression analyses indicate a relation between subjects' initial reactions to the candidates and subjects' subsequent judgments. Generally, the more positive and fewer negative comments subjects made initially, the more likely a candidate was to be hired and the more favorable the evaluation.

TABLE 3
RELATION OF INITIAL REACTIONS TO THE CANDIDATES TO HIRING DECISIONS AND QUALITY RATINGS

| Condition | Frequency of positive comments | | Frequency of negative comments | |
|-----------------|--------------------------------|---------|--------------------------------|---------|
| | Experimental | Control | Experimental | Control |
| Initial hiring | .43* | | -.34 | |
| Quality ratings | .49* | .41* | -.30 | -.49* |
| Second hiring | .40* | | -.20 | |

Notes. (1) $*p < .05$. (2) Entries are beta coefficients from regression equations. The correlations used for the regression analyses were means calculated by converting the correlation for each job applicant to a Z score, using Fisher's r to Z transformation. The Z scores were then averaged for the 12 applicants using the formula described in Hays (1981, p. 467). The resulting mean Z score was converted back to r . (3) The minus signs indicate that the fewer the negative comments, the more likely a person was to be hired, and the higher his quality rating.

Intersubject Agreement on Hiring Decisions

Although it was not a major focus of the current study, we examined the degree to which subjects agreed on the persons to be hired. We calculated agreement using Kappa (K), a statistic that shows the proportion of agreement remaining after chance agreement is removed (Fleiss, 1971; Watkins, 1979). The Kappa statistic can range from negative values (below chance agreement) through zero (chance agreement) to +1.00 (perfect agreement), and is distributed as a standard normal variate.

First, let us consider the initial hiring decisions. In the condition in which 2 candidates were hired, a K of .07 was produced, which is not significantly different from chance agreement. When 6 candidates were hired, a K of .23 resulted, $Z = 2.15$, $p < .05$. This figure indicates that 23% greater than chance agreement was obtained. Though the agreement was significantly above chance, it was not very high in absolute terms.

Intersubject agreement did not rise when subjects were "released" from their initial hiring constraints and were permitted to hire more candidates. The Kappa for the second hiring decision was .076 in the condition in which 2 candidates were hired initially, and .19 in the condition in which 6 candidates were hired initially. Neither value significantly exceeds chance agreement.

The low agreement may be due, in part, to our procedure. We presented subjects with 12 strong candidates. If we had included applicants who were obviously weak, overall intersubject agreement might increase. Subjects would be likely to agree that such mediocre candidates should be excluded from the final pool.

Note, though, that low agreement among judges is not unique to our experimental situation. For example, several studies of agreement among reviewers of manuscripts for major psychology journals obtained Ks of below .30 (Peters & Ceci, 1982; Watkins, 1979).

DISCUSSION

As hypothesized, subjects' evaluation strategies differed as a function of the number of appointments available. The evidence is not that subjects in the 2-appointment condition searched for more faults, but rather that they overlooked some of the positive features of the candidates noted by subjects in the 6-appointment condition.

The obtained interaction is compatible with our theoretical analysis. Allocators in both scarce and relatively abundant resource conditions are motivated to look for faults in applicants. When the resource is abundant, however, subjects have an additional impetus to search for the merits of a candidate. In the current study, subjects in the 6-appointment condition need be just as concerned with finding reasons to accept a candidate, as with finding grounds to reject him.

This asymmetrical impact of scarcity on evaluation strategies is also intuitively appealing. Consider: Your manuscript is rejected by a journal editor. The reviewers' complaints may not surprise you. The flaws are there and you know what they are. What is disturbing, though, is that the reviewers failed to see a number of the positive aspects of the piece which, in your view, outweighed the failings. Allocators are more inclined to accentuate the positive when resources are abundant.

The lack of difference between the initial reactions of control and experimental subjects suggests that the important determinant of evaluations at the predecisional stage is scarcity of the resource, rather than the responsibility of making a choice. In contrast, the quality ratings reflect the impact of having to make a decision. Experimental subjects evaluated the applicants they employed more favorably than control subjects evaluated their own top applicants.⁴

Experimental subjects did not appear to derogate unsuccessful applicants, relative to ratings provided by control subjects. A similar result was obtained by Gerard and White (1983) in a quite different context. In an experiment in which subjects selected paintings, these authors found reevaluation of only the chosen alternatives.

Though this finding does not seem limited to the current experimental context, the possible impact of the simulation aspect of our study must be acknowledged. Because it was a simulation, subjects need not worry about actually harming their rejected applicants. Consequently, there would be little reason for a guilt-induced shift in evaluation of the rejected candidates. Even in a real competition, however, one could suppose that if a rejected applicant is any good, then he or she will get other jobs. In such situations, then, decision makers may be more concerned with justifying their selections than their rejections.

To this point, we have assumed that subjects in the experimental conditions exaggerate the quality of the candidates they hire. Conceivably, the locus of the effect is in the control condition. The quality rating possibly took on greater significance for control than for experimental subjects. Subjects in the experimental conditions presumably saw the hiring decisions as the focus of the study; those in the control condition may have attributed similar significance to the quality ratings. As a result,

⁴ We have reasoned that differences in the quality ratings of experimental and control subjects reflect postdecisional reevaluations by experimental subjects. Conceivably, however, comparable differences may have emerged prior to the hiring decision. We attempted to evaluate this alternative by examining subjects' initial reactions to the candidates (i.e., the number of positive and negative comments). There were no differences between the initial reactions of experimental subjects to candidates they eventually employed in the first hiring session and the initial reactions of control subjects to comparably ranked candidates (top 2 or 6). There were also no experimental-control differences for candidates eventually rejected or ranked low. Although comparing across different measures is hazardous, these data bolster our position that the effects obtained subsequently on the quality ratings emerged postdecisionally.

control subjects' quality ratings may have reflected a concern for impression management that was absent in the experimental condition. Control subjects may have believed that they could present a more perspicacious image of themselves by providing moderate evaluations, rather than appearing too enthusiastic about a candidate.

There is reason to suspect, though, that control subjects' quality ratings are not simply efforts at self-presentation. Control subjects were not more cautious in their initial, tape-recorded evaluations than experimental subjects. Further, these initial evaluations predicted the quality ratings as strongly in the control as in the experimental conditions. Thus, the ratings of control subjects seem to reflect genuine assessments of the candidates.

The second hiring decision, where subjects were given the opportunity to employ as many applicants as they wished, constituted the final dependent measure. Ideally, we would have found little or no increase in number of candidates employed from the first decision. This would have provided the most dramatic evidence that judgments of relative worthiness are converted to judgments of absolute worthiness. Practical considerations led us to construct a set of vitae that made such results highly improbable, however. We had to provide enough good candidates so that subjects in the 6-appointment condition could find 6 desirable applicants and even hire more when the restraints were removed. This meant that the subjects in the 2-appointment condition, in particular, had a number of candidates they could still hire who would grace virtually any department in North America. Nonetheless, those who initially hired only 2 applicants continued to employ fewer candidates than those who originally hired 6.

Tversky and Kahneman (1974) noted that people often fail to make necessary adjustments of initial judgments. For example, in one study subjects were asked to adjust an arbitrary initial estimate of the percentage of African countries in the United Nations. Those starting with "anchors" of 10 and 65% produced adjusted estimates of 25 and 45%, respectively. In the present study, the continued impact of the manipulation of 2 vs. 6 appointments on the number of candidates hired when constraints were removed can be depicted as a failure to adjust an initial "anchor" sufficiently. More important, though, we have described and assessed two psychological processes that appear to mediate this phenomenon, subjects' initial evaluation strategies and their postdecisional shifts in attitude. The implication is that allocators accepted the validity of the results of the winnowing procedure and failed to appreciate fully the impact of external constraints on this process.

Finally, we consider the implications of the current data for the review process in psychology. In discussing the high rejection rates in major psychology journals, Robert Adair (1982), a Yale physicist, noted that such difficulties do not exist in his discipline. He attributed the discrepancy to the "relative simplicity and objectivity of physics and the complexity and subjectivity of many areas in psychology" (p. 196).

Adair subsequently noted, in passing, that the most important physics journal, *Physical Review*, publishes 30,000 pages per year and has an 80% acceptance rate. We suggest that the perception of the relative subjectivity of our science might largely disappear if the *Journal of Experimental Social Psychology* could publish, say, 20,000 pages a year and sustain an acceptance rate of 80%.

Such an expansion may be neither feasible nor desirable. What is important, though, is that social psychologists and others recognize the impact of arbitrarily established high rejection rates on the allocation and inference process. Moreover, social psychologists should not be misled by the apparent facility and efficiency with which the review process is conducted. One troubling aspect of the current data is that subjects see decisions made under scarcity as quite easy to make. As our agreement statistic indicates, however, that ease should not be interpreted as evidence for the reliability of the decision process.

REFERENCES

- Adair, R. K. (1982). A physics editor comments on Peters and Ceci's review study. *The Behavioral and Brain Sciences*, *5*, 196.
- Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 4, pp. 1-34). New York: Academic Press.
- Beyer, J. M. (1978). Editorial policies and practices among leading journals in four scientific fields. *The Sociological Quarterly*, *19*, 68-88.
- Calabresi, G., & Bobbitt, P. (1978). *Tragic choices*. New York: Norton.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford Univ. Press.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378-382.
- Gerard, H. B., & White, G. L. (1983). Post-decisional reevaluation of choice alternatives. *Personality and Social Psychology Bulletin*, *9*, 365-369.
- Hays, W. L. (1981). *Statistics*. New York: Holt, Rinehart & Winston.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. New York: Plenum.
- Myers, J. L. (1979). *Fundamentals of experimental design*. Boston: Allyn & Bacon.
- Olson, J. M., & Zanna, M. P. (1982). Repression-sensitization differences in responses to a decision. *Journal of Personality*, *50*, 46-57.
- Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, *5*, 187-195.
- Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.
- Watkins, M. W. (1979). Chance and interrater agreement on manuscripts. *American Psychologist*, *34*, 796-798.
- Wicklund, R. A., & Brehm, J. W. (1976). *Perspectives on cognitive dissonance*. Hillsdale, NJ: Erlbaum.