COGNITIVE PSYCHOLOGY 26, 103–133 (1994)

The Easy Path from Many to Much: The Numerosity Heuristic

BRETT W. PELHAM, TIN TIN SUMARTA, AND LAURA MYASKOVSKY

University of California, Los Angeles

People are especially sensitive to numerosity as a cue for judging quantity or probability. That is, people sometimes judge amount or likelihood on the basis of the *number of units* into which a stimulus is divided without fully considering other important variables (e.g., the size of the units). People appear to be especially likely to make use of this "numerosity heuristic" when their cognitive resources are taxed. Consistent with this idea, five experiments showed that people are especially likely to overinfer quantity or probability from numerosity (a) when they are asked to make inherently difficult judgments, (b) when they are asked to render judgments while performing a concurrent task, and (c) when they are forced to make especially rapid judgments. In addition to their implications for the numerosity heuristic, the broad implications of these findings for the study of judgment are discussed.

More than half a century ago, in a study of operant conditioning in chickens, Wolfe and Kaplon (1941) found that a kernel of corn that had been divided into four separate pieces served as a more effective reinforcer than a single intact kernel. Although there are several possible explanations for this finding, it is clear, at a descriptive level, that Wolfe and Kaplon's subjects responded not only to the amount of reinforcement received but also to the *number of units* into which the reinforcement was divided. Katz (1930; cited in Hartmann, 1935) reported a similar finding. Chickens became satiated especially quickly, and thus ate less than usual, when their food was divided into multiple units.

Perhaps it is not surprising that a species so low on the phylogenetic pecking order could confuse numerosity for quantity. However, it ap-

This research and the preparation of this manuscript were supported by a seed grant to Brett W. Pelham from the University of California, Los Angeles. We are indebted to Rochelle Gelman, Bob Josephs, Nancy Kanwisher, Bill Swann, Shelley Taylor, Margaret Walsh, and the three reviewers of this paper for their insightful comments on this work; to Mindy Hightower, Jeff Miles, Chris Reyna, Stacey Sinclair, and Lisa Weingarten for their assistance in pilot studies; and to the many students who assisted in the collection of the data for Experiment 4. Finally, we offer numerous thanks to Dan Gilbert. From beginning to end, this work has benefited in countless ways from his generous and continuous advice and assistance. Correspondence and reprint requests should be addressed to Brett W. Pelham, UCLA Department of Psychology, 1285 Franz Hall, 405 Hilgard Avenue, Los Angeles, CA 90024-1563. E-mail should be sent to IYJ6BWP@UCLAMVS.BITNET.

0010-0285/94 $9.00

pears that animals more sophisticated than chickens are also overly sensitive to numerosity as a cue for judging quantity. Traupmann (1971) found that rats acquire an instrumental running response more quickly when rewarded with divided rather than intact food pellets, and Capaldi, Miller, and Alptekin (1989) found that rats prefer four 75-mg food pellets over a single 300-mg pellet. Because *numerosity* (the number of units into which a stimulus is divided) and *quantity* (the total area, weight, or amount of a stimulus) are likely to be strongly correlated in the natural environment of most organisms, inferring quantity directly from numerosity may prove to have benefits that outweigh its potential costs.

DO PEOPLE OVERINFER QUANTITY FROM NUMEROSITY?

There is indirect evidence that people as well as animals sometimes overinfer quantity from numerosity. In studies of group structure and conformity, for instance, Wilder (1977, 1978) demonstrated that when people are encouraged to think of an aggregate of others as several distinct persons rather than a single group, they are more likely to conform to the judgments and opinions of these others (in much the same way that people are more likely to conform to a request made by several people rather than by one person). That is, Wilder (1977) showed that people are more strongly influenced by the behavior of others when these others are divided into a greater number of social units. Further evidence that people may sometimes rely on a "numerosity heuristic" in their judgments can be found in research on attitude change. Specifically, Petty and Cacioppo (1984; Study 1) showed that the same persuasive appeals will sometimes have more impact on attitude change when recipients are simply told (incorrectly) that the appeals contain nine rather than three arguments in support of a particular position.

There is even evidence that people's attitudes about themselves may be influenced by the tendency to overinfer quantity from numerosity. It is well established that possessing a great number of positive specific self-views contributes to a person's overall feelings of self-worth (e.g., see Pelham & Swann, 1989). But like many other abstract concepts, people's specific self-views may be cognitively organized in a variety of different ways, and it appears that people who divide their positive self-views into a greater number of distinct units may be able to boost their overall feelings of self-worth by so doing. Showers (1992) demonstrated that people's global feelings of self-worth were related not only to the content but also to the structure of their specific self-views. More specifically, Showers showed that compartmentalizing a single specific self-view ("I am creative, verbal, and analytical.") into several distinct beliefs ("I am creative. I am verbal. I am analytical.") is associated with especially high self-esteem. Thus, there is indirect evidence that in areas as diverse as conformity, persuasion, and self-evaluation, the numerosity of a stimulus may play a disproportionate role in human judgment.

WHY DO PEOPLE OVERINFER QUANTITY FROM NUMEROSITY?

There are at least three potential reasons that people might overinfer quantity from numerosity. First of all, numerosity and quantity are likely to be as highly correlated in the natural environments of people as they are in the natural environments of animals. From an adaptive perspective, that is, more *pieces* of something usually turns out to be *more* of that something. An eight room house is usually larger than a five room house, two research assistants are usually more productive than one, and three headaches are usually more painful than two. Of course, it is clear that numerosity is not perfectly correlated with quantity. A house of eight small rooms may be much smaller than a house of five spacious rooms, and two indolent research assistants may prove to be nothing but a constant headache. To cite a more straightforward example, slicing the same pizza into eight rather than six pieces obviously does not increase its nutritional value. Thus, just as memorial availability is a useful but imperfect predictor of event frequency, and just as entities that share features with one another are often but not always members of the same category, numerosity may be a useful but imperfect predictor of quantity (e.g., see Kahneman & Tversky, 1973; and Tversky & Kahneman, 1973, 1974, for related discussions of the adaptive underpinnings of the availability and representativeness heuristics; and see Arkes, 1991; Fischhoff, 1982; Fiske & Taylor, 1991; and Simon, 1957, 1981; for discussions of the adaptive aspects of inferential heuristics in general).

Consistent with an adaptive perspective on inferential processes, we assume that most human societies have developed formal number systems precisely because numbers and counting serve a variety of very useful purposes. At the same time, we also assume that counting, like all other judgmental strategies, is likely to have costs as well as benefits. If the numerosity of a stimulus is processed routinely and relatively effortlessly in human judgment (e.g., see Hasher & Zacks, 1979), people might sometimes rely on numerosity as a cue for quantity even in situations in which such a cue could be misleading. Thus, from an adaptive perspective, it is feasible that rats and chickens are not alone in their tendency to overinfer quantity from numerosity.

There is a second theoretical reason that people might overestimate the total quantity or area of a divided as compared with an undivided stimulus. In an extension of a basic psychophysical power function described by researchers such as Stevens (1957) and Kahneman and Tversky (1979), Thaler (1985) pointed out that the relation between the objective magnitude of a stimulus (e.g., the mass of a weight, the volume of a tone) and a person's psychological response to the stimulus (e.g., the perceived heaviness of a weight, the judged loudness of a tone) is not perfectly linear. In particular, Thaler noted that this relation is typically asymptotic; people are very sensitive to small changes near the center or reference point of a range of stimuli but are much less sensitive to equivalent changes at more extreme ranges. As a concrete example, people may perceive the difference between a 1-in. and a 2-in. line to be much greater than the difference between a 5-in. and a 6-in. line. In an application of this principle to marketing and consumer choice, Thaler reasoned that people might sometimes be more satisfied with gains when the gains have been segregated into two or more separate units. This should be true, Thaler argued, because the total judgmental impact of the segregated pieces will typically be greater than the judgmental impact of the whole. As a test of this hypothesis, he asked participants who would be happier: (a) a person who won $50 in one lottery and $25 in another lottery or (b) a person who won $75 in a single lottery. Consistent with the idea that segregating gains makes them more desirable, most of Thaler's participants reported that the first person would be happier (see also Poulton's, 1982, discussion of the *stimulus contraction bias*—the pervasive tendency to overestimate small stimuli and underestimate large stimuli).

The developmental literature on the cognitive abilities of children provides a third theoretical reason that numerosity might sometimes be confused with quantity. This literature has shown that children frequently mistake fundamental cues such as numerosity or stimulus dispersion for quantity (Gelman & Baillergeon, 1983); it has also suggested that even very young children possess a preverbal counting system. Some researchers have even argued that this preverbal counting system exists independent of any specific training in or practice with counting. Presumably, this preverbal counting system plays an important role in a wide variety of judgments, and it may predispose people to be especially sensitive to numerosity as a cue for judging quantity (Gallistel & Gelman, 1991; Gelman, Cohen, & Hartnett, 1989; Gelman & Meck, in press).

Although few adults are as easy to fool as preoperational children (e.g., few adults would fail a conservation task in which they were allowed to watch an experimenter physically divide a stimulus), adults' oversensitivity to numerosity may still make them susceptible to certain forms of nonconservation. For example, if adults are not allowed the luxury of actually observing the division of a stimulus before being asked to assess the total area or quantity of the stimulus, they may show evidence of nonconservation that would not be revealed in a typical Piagetian paradigm.

Although these three alternative arguments for the existence of a numerosity heuristic differ in a number of important ways, they all share at least one important feature. All three suggest that the tendency to overinfer quantity from numerosity may be relatively effortless and thus difficult to resist. At the very least, an adaptive analysis suggests that the tendency is highly overlearned. At the very most, a psychophysical analysis suggests that such a tendency may be a byproduct of perceptual processes that are virtually impossible to bypass—even in adults (see Arkes's, 1991, discussion of the relative difficulty of debiasing psychophysically based judgmental errors, and see Fischhoff's, 1982, discussion of the difficulty of debiasing the overconfidence effect, a judgmental bias that is likely to have a psychophysical basis).[1]

WHEN DO PEOPLE OVERINFER QUANTITY FROM NUMEROSITY?

We have suggested three reasons to believe that people may sometimes overinfer quantity from numerosity. Unlike rats, chickens, and young children, however, most adult judges possess higher-order cognitive capacities that could presumably allow them to resist the temptation to infer magnitude directly from multitude. In other words, when placed in situations that allow them to maximize their judgmental capacities, people may make use of highly sophisticated strategies and produce judgments that are not unduly biased by numerosity. Consistent with this idea, we propose that the degree to which people will overinfer quantity from numerosity will depend, in large part, on the degree to which their higher-order cognitive resources are taxed at the time that they render their judgments. Although we know of no research on judgmental heuristics that has systematically examined this hypothesis, this basic idea is not new. A variety of dual process models of social judgment suggest that people are capable of engaging in systematic as well as simplistic forms of information processing (e.g., see Brewer, 1988; Chaiken, 1987; Chaiken, Liberman, & Eagly, 1989; Fiske & Neuberg, 1990; Gilbert, 1989, 1991, 1993; Ginossar & Trope, 1987; Payne, Bettman, & Johnson, 1992; Petty & Cacioppo, 1986; Trope, 1986).

To select only a few examples, Payne, Bettman, and Johnson (1992) have identified task difficulty or *task complexity* as one of the contextual determinants of whether people engage in "analytical" (i.e., sophisticated) or "intuitive" judgmental strategies. Presumably, the cognitive demands of particularly difficult or complex tasks often force people to

[1] There is a small body of work examining the psychological mechanisms underlying the perception of numerousness (see Birnbaum & Veit, 1973; Ginsburg, 1978; Kaufman, Lord, Reese, & Volkmann, 1949; Taves, 1941), but we know of no previous studies that have systematically examined the role of numerosity in the judgment of area, quantity, or probability.

rely on intuitive rather than analytical processing strategies. Because most problems may be framed in either very friendly or very complex ways, this analysis suggests that different instantiations of the same problem may be handled very differently if one instantiation is easier to think about than the other. Notice that Payne et al.'s analysis suggests that analytical judgment strategies are more "disruptable" than intuitive strategies. According to Payne et al., people will only engage in analytical judgment strategies when their higher-order cognitive resources are not usurped by the excessive demands of a complex or demanding problem.

In his dual process model of causal attribution, Gilbert (1989, 1991, 1993) has made a similar point by arguing that systematic components of causal attribution are typically more disruptable than simplistic or "automatic" components. Gilbert, Pelham, and Krull (1988), in particular, showed that people make use of normative (i.e., systematic) attributional principles such as augmenting and discounting only when they have access to the cognitive resources required to apply these higher order principles. Judges who are cognitively preoccupied with a secondary task (e.g., those who are rehearsing a series of word strings) are likely to draw dispositional inferences about a target ("Geez, what a nervous person!") even in the face of obvious situational constraints on the target's behavior ("And extremely generous, too! I wonder why that guy in the ski mask kept threatening to shoot him?"). Thus, Gilbert's model suggests that the degree to which people will engage in systematic information processing depends on the degree to which their cognitive resources are taxed by the demands of *extraneous, resource-consuming tasks.*

Along similar lines, one of the predictions of Chaiken's (1987; Chaiken, Liberman, & Eagly, 1989; Maheswaran & Chaiken, 1991) heuristic-systematic model of social judgment is that people who are forced to make a judgment or evaluation especially *quickly* may be particularly likely to rely on "heuristic" (i.e., simplistic) rather than "systematic" (i.e., carefully considered) reasoning strategies. Thus, even if people are asked to make relatively simple judgments, and even if people are not cognitively preoccupied with an extraneous task while making these simple judgments, they may fail to engage in systematic processing whenever they are in a hurry to make a judgment (see also Ben Zur & Breznitz, 1981; Petty & Cacioppo, 1986; Rothstein, 1986).

Dual process models such as these typically assume that most judgments can be broken down into relatively effortless and relatively effortful components. Although we realize that the distinction between effortful and effortless information processing is one of degree rather than kind, we concur with the proponents of numerous dual process models of judgment who all assume that it is possible to decompose relatively effortless and relatively effortful judgmental strategies by experimentally manipulating the degree to which people can take advantage of their higher-order cognitive resources. Because relatively effortless processes require little in the way of cognitive resources, reducing people's cognitive resources should interfere very little with such processes (Baddeley & Hitch, 1974; Gilbert, 1989; Gilbert, Pelham, & Krull, 1988; Martin, Seta, & Crelia, 1990; Norman & Bobrow, 1975; Payne, Bettman, & Johnson, 1990; Shiffrin & Schneider, 1977).

In keeping with this assumption, we attempt to build on dual process models of social judgment, on Arkes's (1991) adaptive perspective on judgmental heuristics, and on Simon's (1957, 1981) "satisficing" principle by providing systematic empirical support for a contextual approach to heuristic judgment. In particular, we argue that taxing people's cognitive resources by (a) presenting people with especially difficult problems, (b) giving people extraneous, resource-consuming tasks, or (c) asking people to render especially quick judgments will increase the likelihood that people will base their judgments of area, quantity or probability on the numerosity heuristic. Each of the studies in this paper examines this basic hypothesis. Experiment 1 was designed to test this hypothesis by examining people's judgments of geometric figures that either were or were not divided into multiple units.

EXPERIMENT 1

Method

Participants

Forty-eight introductory psychology students at the University of California, Los Angeles (UCLA) participated in this study for course credit. Participants took part in a mass testing session but were separated from one another by at least three seats to ensure independence in their responses.

Materials and Procedure

Participants estimated the area of two geometric figures—(a) a single intact circle and (b) an identical circle that had been divided into nine separate, wedge-shaped pieces. Independent of this numerosity manipulation, we also manipulated the difficulty of mentally reassembling the divided circle. More specifically, we randomly assigned participants to either a "difficult reassembly" or an "easy reassembly" condition. In the difficult reassembly condition, the nine pieces of the divided circle were arranged along a horizontal line, and in the easy reassembly condition, the nine pieces of the divided circle were arranged in a way that made it obvious that they could be positioned together to form a single circle.

Participants were presented with the stimuli on a single page. Following some instructions at the top of the page, the intact circle appeared about halfway down the page. A triangle (which served as a 1-unit frame of reference for participants' judgments) appeared about ¾ in. below the intact circle, and the divided circle appeared about ¾ in. below the triangle. Participants reported their estimates of the area of the intact circle and the total area of the nine pieces of the divided circle at the bottom of the page. Participants were informed that their answers could take on any value and were asked to use decimal notation to indicate fractional units. The stimuli used in the study appear in Fig. 1.

Results and Discussion

We expected that participants would overestimate the area of a circle if it was divided into many separate pieces. We also expected that this would be especially true when it was difficult for participants to see how the pieces of the circle fit together to form a whole. Both of these predictions were confirmed. In particular, a 2 (numerosity: one unit versus nine, a within-subjects variable) × 2 (difficulty of reassembly: easy versus hard, a between-subjects variable) mixed model analysis of variance (ANOVA) revealed both a main effect of numerosity, $F(1,46) = 21.83$, $p < .001$, and a reliable Numerosity × Difficulty interaction, $F(1,46) = 10.78$, $p = .002$.

Simple effects tests (one-way ANOVAs) clarified the nature of this interaction. First, as illustrated in Table 1, there was only a weak tendency for participants in the easy reassembly condition to perceive the divided circle as larger than the intact circle (surprisingly, however, even this effect approached significance, $F(1,23) = 4.05$, $p = .056$). In contrast, participants in the difficult reassembly condition judged the sum of

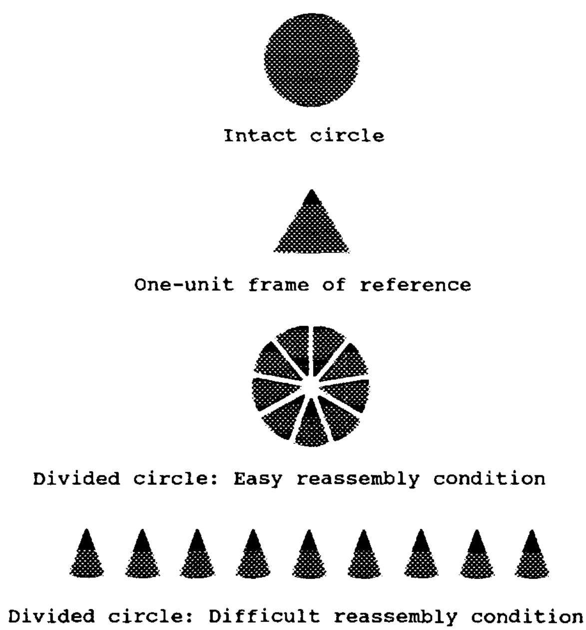

FIG. 1. Stimuli used in Experiment 1.

TABLE 1
Area Estimates of Intact and Divided Circles

Stimulus division	Difficulty of reassembly: Easy	Difficulty of reassembly: Difficult
Intact circle	1.85	1.85
Divided circle	1.99	2.68

Note. Higher values indicate greater judged area (the actual area of the circle was approximately 2.3 units). There were 24 participants in both the difficult and the easy reassembly conditions.

the pieces of the divided circle to be almost 50% larger than the intact circle of the same area, $F(1,23) = 17.96$, $p < .001$.[2]

To summarize, Experiment 1 provides preliminary support for the existence of a numerosity heuristic, and it also identifies one of the boundary conditions of this phenomenon. People seem to be most likely to rely on the numerosity heuristic when they are faced with inherently difficult or ambiguous judgments—that is, under conditions of uncertainty (Tversky & Kahneman, 1974; see also Payne et al.'s, 1992, discussion of decision complexity and Russo's, 1977, analysis of "processability"). One surprising finding of Experiment 1 is that even when we designed our stimuli to promote the use of a normative decision rule, participants still showed some evidence of using the numerosity heuristic. Finally, whereas this study suggests that numerosity may have played an important role in participants' judgments, there is no direct evidence that this was the case. Although our personal experience with pilot participants revealed that some of them counted the divided stimuli (i.e., some participants moved their fingers across the stimuli and paused briefly to touch each element, and one pilot participant, an esteemed member of our faculty, even counted audibly), it is still feasible that our findings were the product of perceptual rather than cognitive processes. In other words, it is possible that we have uncovered a visual illusion rather than a judg-

[2] One reviewer noted that the results of Experiment 1 may be influenced by a potential tendency for people to judge some geometric shapes differently than others. More specifically, if people generally *overestimate* the area of triangles (or wedges) and *underestimate* the area of circles, then the results in the difficult reassembly condition could be misleading. Because we knew of no literature addressing this potential confound, we conducted a brief follow-up study to address this alternative explanation. This study revealed a very robust numerosity effect when nine very small circles were substituted in place of the nine wedges used in Experiment 1. Not surprisingly, however, we could not come up with a way to debias people's judgments in this case because no matter how the nine circles were arranged, it remained difficult for participants to see how the total area of the nine circles compared with the total area of the large circle.

mental bias grounded in numerosity (cf., however, Neisser, 1976, and Newtson, 1980, who argue that the distinction between perception and inference has often been overstated). In keeping with this concern, Experiment 2 was designed to examine the role of numerosity in judgments of more abstract stimuli.

EXPERIMENT 2

Method

Overview

Participants took part in an experiment in which they attempted to solve addition problems presented on a video monitor. Half of the participants were given only this arithmetic task, and half were given a concurrent digit detection task. We expected that relative to one-task participants, two-task (i.e., cognitively taxed) participants would be especially likely to conclude that problems with more elements had larger sums, even when the many element problems contained many *small* elements.

Participants

Twenty-eight UCLA introductory psychology students took part in the experiment for course credit. One participant (from the one-task condition) failed to respond on some of the trials and was replaced by an additional participant.

Procedure

Participants attempted to solve 24 addition problems that appeared in rapid succession on a video monitor. Each problem appeared for 2000 ms and was followed by a 7000-ms pause (a blank screen without a mask), during which time participants were instructed to announce their answers to the experimenter (who immediately recorded them). Twelve problems were designated "few-element" problems (mean number of elements per problem = 4.75) and 12 problems were designated "many-element" problems (mean number of elements per problem = 8.41). Because each few-element problem (e.g., "7.7 + 12.0 + 6.2 + 8.1") was identical in value to one of the many-element problems (e.g., "3.6 + 5.3 + 6.5 + 10.2 + 2.1 + 3.7 + 1.8 + 0.8"), the number of elements in a problem was independent of the correct answer to the problem.

The 24 problems were presented in random order with the constraint that few-element problems and many-element problems having the same solution (i.e., same-solution pairs) were always presented consecutively. In addition, for 6 of the 12 pairs of problems, the few-element problem was presented first, and for the other 6 pairs, the many-element problem was presented first. Before participants began the addition task, they were informed that the correct answer to each problem was a whole number, and they were instructed to guess if they were uncertain of their answers. For purposes of later analysis, we computed composite (average) measures of both participants' answers to the 12 few-element problems and their answers to the 12 many-element problems. Participants randomly assigned to the one-task condition performed only the addition task described above. Participants randomly assigned to the two-task condition performed a digit-detection task in addition to the addition task.

The digit-detection task. Before beginning the addition task, participants in the one-task condition were told that the rapid presentation of the problems would make the experiment especially demanding. Participants in the two-task condition were told that the rapid presentation of the problems, combined with the requirements of a second task, would make the experiment especially demanding. This second task, the digit-detection task, consisted of scanning each problem for the presence of the digit "4" (which appeared in 14 of the 24 problems) and providing a yes or no response to the digit-detection task before reporting the answer to the addition problem. As a manipulation of cognitive load, the digit-detection task had several desirable features. First, the digit-detection task guaranteed that participants would be cognitively preoccupied during the presentation of *each* of the 24 individual addition problems. Second, instead of drawing participants' attention away from the addition problems, the digit-detection task explicitly required participants to attend to the addition problems themselves. Without exception, two-task participants who commented on the digit-detection task after the experiment reported that it was highly demanding, and preliminary analyses indicated that, as a group, two-task participants performed well, but not perfectly, on the digit-detection task (mean number of correct responses = 19.9; range = 16–24).

Results and Discussion

We expected that cognitively loaded participants would be especially likely to infer quantity directly from numerosity. This prediction was confirmed. In particular, a 2 (numerosity: few-element versus many-element problems) × 2 (cognitive load: high versus low) mixed model ANOVA revealed, in addition to a main effect of numerosity, $F(1,26) = 7.48$, $p = .011$, an interaction between numerosity and cognitive load, $F(1,26) = 5.33$, $p = .029$. As illustrated in Table 2, one-task participants correctly assigned similar values to the few-element and the many-element problems, $F < 1.0$. In contrast, two-task (i.e., cognitively taxed) participants judged the many-element problems to be larger than the few-element problems, $F(1,13) = 12.61$, $p = .004$. From a different perspective, one- and two-task participants gave similar answers to the (easier) few-element problems, $F < 1.0$, but in the case of the (more difficult) many-element problems, two-task participants gave larger answers, $F(1,26) = 4.19$, $p = .051$. Like the findings of Experiment 1, these findings suggest that people are especially likely to engage in heuristic processing when faced with especially difficult judgments. In addition, they provide converging evidence that the use of the numerosity heuristic is moderated by cognitive load.

TABLE 2
Estimated Solutions to Addition Problems

	Cognitive load	
Numerosity	One task	Two tasks
Few-element problems	37.3	38.7
Many-element problems	37.9	45.9

Note. Higher values indicate larger answers (the actual mean of the 24 addition problems was 42.5). There were 14 participants in both the one-task and the two-task conditions.

Evidence of Anchoring and Adjustment

In addition to the numerosity heuristic, participants in Experiment 2 may have made use of a better known judgmental heuristic. In particular, participants may have employed an *anchor and adjust* strategy in their judgments by estimating the sum of the first few elements in each addition problem and then adjusting these inferences (insufficiently) for the remaining elements (e.g., see Tversky & Kahneman's, 1974, "factorial" problem). Unlike an overreliance on numerosity, however, a reliance on an anchor and adjust strategy should cause participants to *overestimate* the value of the few-element problems relative to the many-element problems. (This should be true because the first few numbers in the many-element problems, like all of the other numbers in these problems, were typically almost twice as large as the first few numbers in the many-element problems.) At an aggregate level, this obviously did not happen (as shown above). It is possible, however, that the numerosity heuristic may have overshadowed the anchor and adjust heuristic, especially in the two-task condition (where participants were required to attend, at least minimally, to all of the elements in each problem).

This appears to have been the case. Consistent with the anchor and adjust heuristic, within-subject correlations (computed across the 24 addition problems for each individual participant) revealed that participants' judgments (i.e., their answers to the 24 problems) were more strongly correlated with the value of the first as opposed to last element in the problems (mean respective within-subject *r*s were .47 and .25). Moreover, this evidence of anchoring and adjustment was less pronounced among two-task participants, tentatively suggesting that the use of numerosity as a cue for judging quantity may be *more* effortless than the use of the anchor and adjust heuristic. In other words, the fact that cognitively loaded participants made (a) great use of the numerosity heuristic and (b) only minimal use of the anchor and adjust heuristic tentatively suggests that the use of numerosity as a cue for judging quantity may be less cognitively demanding than the use of an anchor and adjust strategy.

Evidence of the Moderating Role of Cognitive Load

The results of Experiment 2 suggest that, unlike the normative process of engaging in mental arithmetic, making use of the numerosity heuristic is a relatively effortless judgmental strategy. Presumably, one-task participants were able to devote a substantial portion of their cognitive resources to the normative, analytical process of adding specific values rather than counting elements. In contrast, participants whose cognitive resources were experimentally depleted produced judgments that were based largely on simple cues for numerosity.

Is there any direct evidence that this was the case? Yes. First of all, at a within-subjects level (i.e., across the 24 problems), the judgments of two-task, as compared with one-task participants, were especially likely to covary with numerosity. More specifically, within-subject correlations between numerosity (the number of elements in a specific problem) and judgment (the answer offered for a specific problem) revealed that, relative to the judgments of one-task participants, the judgments of two-task participants were more strongly associated with the number of elements in the problems (mean respective within-subject *r*s were .18 and .38, $F(1,26) = 5.53$, $p = .027$ for a between-subjects comparison between these within-subjects *r*s).

Focusing solely on the two-task participants, there was further evidence that increased cognitive load was associated with increased use of the numerosity heuristic. In particular, the two-task participants who made the greatest number of correct responses on the digit-detection task (and thus those who presumably were the most cognitively taxed) also showed the greatest tendency to report that the many-element problems were larger than the few-element problems, $r(12) = .62$, $p = .019$. The more two-task participants were cognitively preoccupied, the more they appeared to use the numerosity heuristic. Together, these supplemental analyses suggest that making judgments on the basis of numerosity is a relatively effortless form of information-processing. Presumably, cognitive load disrupts normative processing strategies, forcing people to rely on more simplistic strategies (see Baddeley & Hitch, 1974; and Gilbert, 1989, for related findings).

Taken together, the findings of Studies 1 and 2 suggest that when people's cognitive resources are taxed (by giving them difficult tasks or by giving them two tasks), they will be unable to make use of higher-order inferential rules and will thus rely disproportionately on numerosity as a cue for inferring quantity. Presumably, this is true because numerosity can be assessed more quickly and easily than more sophisticated cues. If people make quick judgments based on numerosity and then correct these preliminary judgments after more sophisticated processing, it should be possible to decompose the estimation of quantity by manipulating the amount of time people are given to render a judgment (see Rothstein, 1986; Shiffrin & Schneider, 1977; Swann, Hixon, Stein-Seroussi, & Gilbert, 1990). If the estimation of quantity from numerosity happens especially quickly, then forcing people to render especially quick judgments should increase their use of the numerosity heuristic. Experiment 3 was designed to test this prediction.

EXPERIMENT 3

Method

Overview

Participants estimated the monetary value of arrays of American coins that appeared on a video monitor. Although all participants were asked to make these judgments relatively quickly, some participants were asked to make their judgments more quickly than others. We expected that those asked to render especially rapid judgments would be especially likely to base their judgments on the number rather than the nature of the coins whose values they judged.

Participants

Twenty-six UCLA introductory psychology students took part in the experiment for course credit. Two participants (one from the brief presentation condition and one from the extended presentation condition) failed to provide judgments for some of the stimuli, and one participant from the extended presentation condition admitted having knowledge of our hypotheses prior to taking part in the study. These three participants were replaced with three new participants.

Procedure

Participants estimated the total values of eight arrays of American coins presented in rapid succession on a 13 × 14-in. video monitor. Participants were seated approximately 30 in. from the video screen, and the coins were presented at eye level at approximately 180% of their actual size. Only nickels, dimes, and quarters were used in the arrays, and in all eight arrays, the coins were displayed face up on a flat black background and arranged arbitrarily.

As in Experiment 2, we manipulated the number of stimuli that participants judged while holding the total value of the stimuli constant across the two numerosity conditions. We did this by creating matched arrays of coins that had the same monetary value but were composed of different numbers of coins. More specifically, four of the arrays were designated "few-coin" arrays (mean number of coins = 8.75) and four of the arrays were designated "many-coin" arrays (mean number of coins = 19.25). For example, one of the few-coin arrays consisted of eight coins totaling $1.30 (four quarters, two dimes, and two nickels). The many-coin array corresponding to this array contained 20 coins having the same total value (one quarter, two dimes, and 17 nickels). The eight arrays were presented in random order with the constraint that the matched pairs of coins were never presented consecutively.

Presentation Rate Manipulation

All participants viewed the same eight arrays of coins in the same random order, and all participants were allowed up to 5000 ms after the presentation of the coins to announce their estimate of the total value of the coins (during this 5000-ms buffer the video screen remained blank, but after 3000 ms the screen changed color to provide a 2000-ms warning before the presentation of the next array). For participants randomly assigned to the extended presentation condition, the stimulus arrays appeared for 5000 ms before the buffer screens appeared. In contrast, in the brief presentation condition, the coins appeared for only 500 ms before the buffer screens appeared (it should be noted, however, that participants in both conditions typically made use of most of the 5000-ms buffer before rendering their judgments).

Results and Discussion

The results of Experiment 3 were consistent with predictions. Specifically, a 2 (numerosity: few- versus many-coin arrays) × 2 (duration of presentation: brief versus extended) mixed model ANOVA revealed, in addition to a main effect of numerosity, $F(1,24) = 9.76$, $p = .005$, an expected Numerosity × Duration interaction, $F(1,24) = 4.72$, $p = .040$. Simple effects tests revealed that this interaction was consistent with predictions. In particular, as shown in Table 3, participants in the extended presentation condition showed only a weak (and nonsignificant) tendency to overestimate the value of the many-coin arrays, $F(1,12) = 1.17$, $p > .30$. In contrast, participants in the brief presentation condition judged the many-coin arrays to have almost twice the value of the few-coin arrays, $F(1,12) = 8.69$, $p = .012$.

Evidence of the Use of Numerosity Cues

Although it is clear that rushed participants overestimated the value of the many-coin arrays, there is no guarantee that these overestimates were grounded in an overreliance on numerosity. A more detailed analysis of participants' responses, however, suggests that this was the case. As illustrated in the lower half of Table 4, within-subject correlations analogous to those computed in Experiment 2 revealed that the judgments of rushed participants were especially likely to covary (across the eight trials of the experiment) with the number of coins in the arrays, $F(1,24) = 5.69$, $p = .025$, for a between-subjects comparison between the mean within-subjects *r*s. Conversely, as illustrated in the upper half of Table 4, the judgments of nonrushed participants were especially likely to covary (across the eight trials) with the total monetary value of the coins, $F(1,24) = 3.81$, $p = .063$. These supplemental analyses suggest that the judgments of rushed participants were grounded in the quantity rather than the quality of the stimuli they judged.

TABLE 3
Value Estimates of Arrays of Coins

	Stimulus presentation duration	
Numerosity	Extended	Brief
Few coins	1.39	1.30
Many coins	1.60	2.46

Note. Higher values indicate larger answers in dollars (the mean dollar value of the eight arrays was 1.425). There were 13 participants in both between-subject conditions.

TABLE 4
Judgment-Value and Judgment-Numerosity Within-subject Correlations

Correlation	Stimulus presentation duration	
	Extended	Brief
Judgment-value	.62	.39
Judgment-numerosity	.17	.54

Note. Higher values indicate larger mean within-subject *r*s (theoretical range = −1 to +1). There were 13 participants in both between-subject conditions and eight predictor and criterion observations per participant.

Despite the evidence of mediating processes presented in Studies 2 and 3, an important limitation of the present findings is that it is still difficult to know that numerosity *per se* was the judgmental cue used by our cognitively taxed participants. As revealed by work on the conservation of number in children (McLaughlin, 1981) and by work in animal cognition (Capaldi & Miller, 1988), other fundamental perceptual cues (e.g., area, density, stimulus dispersion) are often confounded with numerosity. Although we have tried to control for these cues wherever possible (e.g., Experiment 1 controlled for area, Experiment 2 controlled for stimulus density) the case for the numerosity heuristic would be strengthened by evidence that numerosity affects judgments that are unrelated to stimulus dispersion, density, or area. One kind of judgment that fits this category is the estimation of probability. Experiment 4 examined the role of the numerosity heuristic in people's probability judgments. More specifically, Experiment 4 assessed people's preferences for desirable or undesirable activities whose probabilistic outcomes either were unrelated or were negatively related to their numerosity.

EXPERIMENT 4

Method

Overview

Participants were presented with two hypothetical scenarios, each of which required them to make a judgment or decision. The first scenario involved a choice between two negative options that differed in numerosity (i.e., crossing one of two mine fields that contained different numbers of mines), and the second scenario involved a choice between two positive options that differed in numerosity (i.e., entering one of two lotteries that provided entrants with different numbers of tickets). In addition to this manipulation of outcome valence, we also manipulated the difficulty of using the normative decision rules required to make these choices correctly. We expected that when we made it easy to use the appropriate normative decision rules, participants would resist the use of the numerosity heuristic and make their decisions rationally. In contrast, we expected that when we taxed participants' cognitive resources by making the options they considered difficult to think about, they would use the numerosity heuristic.

Participants

One hundred thirty UCLA undergraduates took part in the experiment on a voluntary basis.

Materials and Procedure

Participants were asked to fill out a brief, anonymous survey on "judgment and intuition." Participants learned, from a set of written instructions, that the survey could contain one or more questions that did not have a correct answer, and they also learned that the questions that did have correct answers could be either very difficult or very easy.

The mine field question. In addition to a distractor item that contained no manipulations, the survey contained two questions that involved numerosity manipulations. The first of these questions asked participants to play the role of a soldier who was forced to cross one of two enemy mine fields, and the crucial portion of the difficult version of the question read as follows:

> One field (Field A) contains 5 mines, all of which must be stepped on to get across the field and each of which has a 20 percent chance of exploding when stepped on. The other field (Field B) contains 10 mines, all of which must also be stepped on and each of which has a 10 percent chance of exploding when stepped on (thus, you will step on twice as many mines in Field B, but each mine you step on will be only half as likely to explode when you step on it). Which escape route would you choose?

Participants expressed their preference for one of the two routes on a 7-point scale that ranged from "WOULD DEFINITELY CHOOSE ROUTE A" to "WOULD DEFINITELY CHOOSE ROUTE B" (the midpoint, 4, was labeled "NO PREFERENCE"). Since the field with 10 mines offers a 35% chance of survival ($.9^{10} = .35$) and the field with five mines offers a 33% chance of survival ($.8^5 = .33$), the normatively correct answer is 7 (the field with 10 mines). Because these calculations are cognitively demanding, however, we expected that participants who received this version of the question would rely on their intuition rather than their knowledge of probability and would express a preference for the route with fewer mines.

The lottery question. The final question in the survey involved a choice between two positive rather than two negative options. The difficult version of this question read as follows:

> Suppose that you are invited to enter one of two lotteries, each having a million entrants, and each offering a single prize of $1 million to the winner. In Lottery A, participants are all given one lottery ticket. In Lottery B, participants are all given ten lottery tickets. Which lottery would you prefer?

The 7-point scale for this question was an appropriately reworded version of the scale for the previous question. Since this question made it clear that *all participants* would receive either 1 or 10 lottery tickets, the normatively correct answer to this question is 4 (no preference). We expected, however, that because participants would be unlikely to consider this base-rate information, they would disproportionately base their preferences on numerosity and rate the 10-ticket lottery as especially desirable.

The difficulty of calculation manipulation. Independent of the numerosity cues provided in the mine field and lottery questions, we also manipulated the ease with which participants could engage in the higher-order calculations required to answer the questions correctly. In both cases, we did so by reducing the computational demands of the questions for some participants. In the mine field question, we did this by telling participants that field A contained only one mine (that had a 100% chance of exploding when traversed) and that field

B contained two mines (each of which had a 50% chance of exploding when traversed). Notice that although this version of the problem is quite similar to the original, it makes it much easier to compute the probability of survival in each field (0% versus 25%). Along similar lines, we made base-rate information especially easy to use in the easy version of the lottery question by telling participants that each lottery was to have "only two entrants (yourself and a stranger named Joe)." For this version of the question, we expected that participants would realize that there was no advantage to possessing 10 lottery tickets when all the other participants (Joe) held the same number of tickets. To simplify the data analysis, difficulty of calculation was manipulated solely on a between-subjects basis—that is, each participant received either two difficult or two easy problems.

Results and Discussion

The results of Experiment 4 were similar to the results of the previous studies. In particular, a 2 (valence of outcome: desirable versus undesirable) × 2 (difficulty of calculation: easy versus hard) mixed model ANOVA revealed, in addition to an uninformative main effect of difficulty of calculation, $F(1,126) = 10.02$, $p = .002$, a reliable Valence × Difficulty interaction, $F(1,126) = 35.51$, $p < .001$.

Simple effects tests revealed that this interaction was consistent with predictions. As revealed in the left-hand column of Table 5, participants in the easy calculation condition expressed preferences that were highly normative. First of all, at a within-subject level, they clearly reported that their interest in many mines was greater than their interest in many lottery tickets, $F(1,63) = 22.39$, $p < .001$. Second, the mean preference in the lottery question was virtually identical to the normative value of 4.0, $F < 1.0$. Third, their interest in the more heavily mined (but safer) of the two mine fields easily exceeded the theoretical midpoint of the scale, $F(1,63) = 42.42$, $p < .001$, for a comparison of the observed mean with the theoretical mean of 4.0.

In contrast to the "rational" answers of participants in the easy calculation condition, the preferences expressed by participants in the difficult calculation condition clearly violated normative decision rules. As illustrated in the right-hand column of Table 5, these cognitively taxed participants not only expressed greater interest in numerous lottery tickets

TABLE 5
Preferences for Positive and Negative Hypothetical Outcomes

Valence of outcome	Difficulty of calculation: Easy	Difficulty of calculation: Hard
Negative	5.63	3.38
Positive	4.03	4.81

Note. Higher values indicate greater preference for the high numerosity option. There were 64 participants in both between-subject conditions.

than in numerous mines ($F(1,63) = 14.22$, $p < .001$), they also expressed both special interest in the lottery with many tickets and special disinterest in the mine field with many mines (both $ps < .05$ for comparisons with the theoretical value 4.0). In other words, these cognitively taxed participants expressed a preference for more of a good thing, even when doing so could do them no good. Similarly, they expressed a preference for the lesser of two evils even when the lesser evil was normatively less preferable.

Alternative Explanations for Experiment 4

Although the findings of Experiment 4 appear to support our hypothesis, there are at least two potential problems with this study. One problem is that in the mine field problem our manipulation of problem difficulty was confounded with the desirability of the options presented. That is, in the easy version of the mine field problem, the difference between the "safe" and the "dangerous" mine field was noticeably larger (a 0% versus a 25% chance of survival) than the difference between these options in the difficult version of the same problem (33% versus 35%). Although this confound cannot explain why participants in the difficult calculation condition preferred 5 mines over 10, it could be partly responsible for our results.

Although the lottery question does not suffer from this problem, it may suffer from a different problem. If our use of the word "lottery" caused some participants to conjure up images of how the *typical* lottery works, they may have ignored the base-rate information provided in the problem, assuming that in this hypothetical lottery, as in most real lotteries, possessing more tickets increases one's chances of winning. Moreover, it is possible that making the problem more vivid and personalized (in the easy calculation condition) may have diminished people's tendencies to make use of such a "lottery script." In a follow-up study patterned closely after the original study, we attempted to address each of these problems.

To address our concerns about the lottery problem, we changed the word "lottery" to "sweepstakes" in a replication study conducted on 102 UCLA undergraduate volunteers.[3] To address our concerns about the mine field problem, we added two control groups in the replication study.

[3] Because of our interest in the numerosity heuristic, we have collected numerous sweepstakes entries over the past 2 years. Without exception *all* of the offers we have received provided us with more than one entry. Thus, if our participants' a priori theories of sweepstakes suggest that 10 entries are better than one, they are based on an erroneous understanding of the rules governing most real sweepstakes. If we assume that sweepstakes promoters use this multiple-entry strategy because they have found it to be more effective than a single-entry strategy, this would provide support for the external validity of our findings in Experiment 4.

More specifically, half of the participants in the replication study were assigned to conditions similar to those of the original study, and half were given only one numerosity problem: one of two highly simplified versions of the mine field problem. In the "large discrepancy" version of the new problem, participants were asked to choose between (a) a single mine that had a 100% chance of exploding and (b) a single mine that had a 75% chance of exploding. In the "small discrepancy" version of the problem, the values 100 and 75 were replaced with the values 67 and 65, respectively.

As illustrated in Table 6, this follow-up study yielded a replication of our original results. Like Experiment 4, this study yielded a highly reliable Valence × Difficulty interaction for the primary analysis, $F(1,51) = 26.50$, $p < .001$. Simple effects test revealed that, in comparison with participants in the easy calculation condition, those in the difficult calculation condition not only reported that they found many sweepstakes tickets especially desirable, $F(1,51) = 7.96$, $p = .007$; they also reported that they found many land mines especially undesirable, $F(1,51) = 24.28$, $p < .001$.

Analyses from the participants in the large and small discrepancy control groups revealed that the differences observed for the mine field problem were not due to any differences in the safety of mines across the two conditions. Participants in the large and small discrepancy control groups reported equally strong preferences for the safer of the two mine fields; respective means were 6.32 and 5.92, $F < 1$. The responses of participants in these two groups were also highly similar to those of participants in the easy calculation condition, both Fs < 1, and highly dissimilar to the responses of participants in the difficult calculation condition, both Fs $>$ 22.2, both ps $< .001$.

Experiment 4 thus suggests that the operation of a numerosity heuristic may play a role in the assessment of probability as well as area or quantity. In their work on "fault trees" Fischhoff, Slovic, and Lichtenstein (1978, Study 5) presented evidence consistent with this position. In par-

TABLE 6
Preferences for Positive and Negative Hypothetical Outcomes

	Difficulty of calculation	
Valence of outcome	Easy	Hard
Negative	5.90	3.78
Positive	4.03	5.35

Note. Higher values indicate greater preference for the high numerosity option. There were 30 participants in the easy calculation condition and 23 participants in the difficult calculation condition.

ticular, these researchers found that dividing a set of potential causes for an event into two separate subsets makes the total set of causes appear more likely. Experiment 4 also provides further evidence that people are most likely to base their judgments on numerosity when their cognitive resources are experimentally taxed. One limitation of Studies 1–4, however, is that they all focus on nonsocial stimuli. Because many of people's important decisions in life involve social stimuli, a judgmental rule of thumb that applied only to nonsocial stimuli would be of limited practical or theoretical utility. In keeping with this consideration, Experiment 5 was designed to examine the role of the numerosity heuristic in social perception.

EXPERIMENT 5

Method

Overview

In a variant of a paradigm developed by Asch (1946; see also Anderson, 1965) participants read one of two lists of nine positive traits describing a hypothetical stranger. In one list, the nine traits were listed separately. In the other list, the same traits were organized into three groups of three traits each. Independent of this numerosity manipulation, participants were given either 1 or 3 min to make a series of global evaluations about the person whose description they read. Consistent with the idea that people who are forced to make rapid judgments will be particularly likely to rely on the numerosity heuristic, we expected that in the nine traits condition, participants given only 1 min to make their judgments would evaluate the target more favorably than those who had more time to consider their judgments. This should be true because participants in the 1-min condition would not have time to engage in all of the systematic comparisons required to reveal that many of the nine positive traits used to describe the target were redundant with one another (and thus did not truly constitute complete units). In contrast, we expected that participants in the 3-min condition would be able to discover the redundancy in the nine traits used to describe the target and would make judgments that reflected this systematic discovery.

Participants

Thirty-six UCLA psychology students took part in the experiment for course credit. One participant failed to follow instructions and was deleted from the study.

Procedure

Participants read a list of nine positive traits describing a hypothetical person named Tom. Participants randomly assigned to the "separate traits" condition read and copied a list of nine separate, numbered traits (1, artistic; 2, athletic; 3, clever; 4, coordinated; 5, creative; 6, energetic; 7, intelligent; 8, musical; 9, smart). Participants randomly assigned to the "grouped traits" condition were presented with the same nine traits organized into three groups of three traits each (1, artistic, creative, musical; 2, athletic, coordinated, energetic; 3, clever, intelligent, smart). Participants were given the copying task to ensure that, at some level, all participants encoded each specific trait. The order of presentation of the traits was counterbalanced across subjects using three different orders in a Latin-square design (in the separate traits condition the traits were blocked into groups of three and these blocks were presented in three different orders). After reporting their impressions of Tom, participants turned the page and were given an unexpected free-recall measure of their memory for the nine traits before being debriefed by the experimenter.

Duration of Evaluative Processing

Before beginning the copying task, participants were informed that after they had copied the last trait in their list, they would be given a limited amount of time to think about the traits presented and to organize their impression of Tom. Participants randomly assigned to the brief consideration condition were given 1 min to think about Tom before turning the page to report their impressions. Those assigned to the extended consideration condition were given 3 min to think about Tom before reporting their impressions. In both conditions, participants were specifically instructed to try to form an overall impression of Tom's level of competence and talent and were encouraged to use all of the time available to organize their impression.

Dependent Measures

Participants reported their impressions of Tom in terms of how *well-rounded, talented,* and *capable* they perceived him to be (e.g., "Compared to the average person, how well-rounded is Tom?"). Participants made each of these global competence ratings on 11-point scales. The endpoints for the first question, for example, were labeled *not at all well-rounded* and *extremely well-rounded.* Immediately following each global competence rating, participants reported the *certainty* of that rating on a scale anchored at the endpoints by (1) *not at all certain* and (11) *extremely certain.* To produce an overall index of the strength of participants' impressions of Tom, each trait score was multiplied by its corresponding certainty score. These three scores were then averaged to produce a single score. Participants' responses to the resulting three-item scale proved to be internally consistent (α = .75). Responses on this index also proved to be closely related to both a composite measure of the three trait scores considered separately and a composite measure of the three certainty scores considered separately; respective *r*s were .82 and .93.

Results and Discussion

The results of Experiment 5 provided further support for our hypothesis. A 2 (organization of traits: grouped versus separate) × 2 (duration of processing: brief versus extended) completely between-subjects ANOVA revealed only an expected Organization × Duration interaction, $F(1,31) = 4.40$, $p = .044$.

As illustrated in Table 7, simple effects tests revealed that this interac-

TABLE 7
Impressions of a Hypothetical Stranger

	Duration of evaluative processing	
Numerosity	Extended	Brief
Grouped traits	75.3	67.1
Individual traits	69.0	88.9

Note. Higher values indicate more positive judgments (theoretical range = 1–121). There were eight participants in the extended presentation/grouped traits condition and nine participants in all other conditions.

tion was consistent with our hypothesis. For participants who were given time to consider the nine traits at great length, the grouping manipulation had no effect. Those who were presented with three groups of three traits viewed Tom just as favorably as those who were presented with a list of nine separate traits, $F < 1$. (Presumably this was true because those in the separate traits condition realized that there was some redundancy in the nine traits.) In contrast, participants who were forced to render snap judgments reported more favorable impressions of Tom when his nine favorable features were listed individually, $F(1,16) = 4.44$, $p = .051$.[4]

These findings are reminiscent of the work of Newtson (1973). Using a more naturalistic design than the one used here, Newtson found that when judges are induced to segment a target's behavior into especially small perceptual units, they make especially confident attributions about the target's behavior. Although this is only one possible interpretation, it is feasible that the participants in Newtson's research who segmented behavior into a great number of small units may have subjectively felt that they had more *information* about the targets they judged. Independent of how we interpret Newtson's work, the work reported here suggests that people make use of the numerosity heuristic in a wide range of judgments and decisions. Independent of the particular content of their judgments, people also appear to be especially sensitive to numerosity as a cue for inferring quantity when they are forced to make decisions without the full benefit of their higher-order inferential abilities.

GENERAL DISCUSSION

The studies discussed in this paper provide converging evidence for the existence of a numerosity heuristic, a judgmental strategy in which people disproportionately base their judgments of area, quantity or probability on the number of units into which a stimulus is divided. Like rats and chickens, people may be so attentive to numerosity that they sometimes show failures of conservation (Capaldi & Miller, 1988; Capaldi, Miller, &

[4] Separate analyses of composites of the favorability and the confidence ratings revealed that only the confidence ratings yielded the predicted interaction. However, planned comparisons on the favorability ratings (separate simple effects tests for the extended and brief evaluation conditions) revealed that the effect of the numerosity manipulation was significant only in the brief evaluation condition; in the extended evaluation condition, numerosity had no effect ($F < 1$). Not surprisingly, participants in the grouped traits condition had significantly better memory for the traits than participants in the separate traits condition (see Miller, 1956). In addition, there was a weak, and nonsignificant, tendency for participants in the extended consideration condition to remember more of the traits. These main effects, however, cannot account for the significant interaction between grouping and duration. In fact, an analysis of covariance (ANCOVA) that equated all groups for memory yielded a slightly stronger pattern of results than those reported in the text.

Alptekin, 1989). At the same time, our findings also indicate that people are capable of resisting the temptations of the numerosity heuristic. The degree to which people will do so, however, appears to be intimately linked to the demands placed on their higher-order inferential abilities.

Because this research suggests that people make use of the numerosity heuristic only when they are cognitively taxed, it is tempting to conclude that the numerosity heuristic is a strategy of last resort—one that people rely on only when their cognitive resources are seriously depleted. Although this is one reasonable interpretation of our findings, an equally reasonable interpretation is that the numerosity heuristic is a strategy of first resort. If inferring quantity or probability directly from numerosity is less cognitively demanding than engaging in more systematic reasoning strategies, it is possible that the use of the numerosity heuristic represents a "default" strategy upon which people rely very heavily in their spontaneous judgments in their daily lives. This analysis is consistent with speculations offered by Arkes (1991), Fiske and Taylor (1991), and Simon (1957, 1981) who all suggest that people rely heavily on overlearned decision rules (i.e., heuristics) not only because the pressing demands of daily life force people to become "cognitive misers" but also because the rapid answers provided by heuristics frequently prove to be correct (see also Gilbert, 1989, 1991, 1993; Stich, 1990).

Although it is difficult to know exactly how often people apply systematic versus heuristic reasoning strategies in their daily lives, we assume that a number of distinct requirements must all be met before people can apply a systematic decision rule. To make a systematic judgment, people must be motivated to make a highly systematic judgment, they must be aware of the correct decision rule, they must encode (or recode) the problem in a clear and manageable form, and they must have both the time and the available cognitive resources required to apply the systematic decision rule (see, for example, Martin et al., 1990). In light of this analysis, we would be surprised indeed if people were able to make all of the quantity and probability judgments that they are forced to make in a typical day without occasionally relying on some form of numerosity heuristic.

The Effortlessness of Numerosity

Although we realize that very few cognitive processes can be considered completely effortless, the idea that the numerosity heuristic represents a satisficing strategy is consistent with the idea that it is a relatively effortless judgmental process. This perspective on the numerosity heuristic is consistent with empirical evidence independent of that reported here. Most notably, research by Chaiken (1987; Chaiken, Liberman, & Eagly, 1989) and by Petty and Cacioppo (1984, 1986) has revealed that simple cues such as argument numerosity play an important role in attitude change only when people engage in "peripheral" or "heuristic" (low effort) rather than "central" or "systematic" (high effort) forms of information processing. Along these lines, Petty and Cacioppo (1984; Study 2) demonstrated that when people find an attitudinal issue uninvolving (a condition that presumably encourages peripheral processing), they show more attitude change in response to a large number of weak arguments than they do in response to a small number of strong arguments. In contrast, when people are motivated to engage in systematic thought about an issue (i.e., when issue relevance is high), people are strongly affected by the nature rather than the number of the arguments they hear. This work, like Chaiken's (1987; Chaiken, Liberman, & Eagly, 1989) heuristic-systematic model of persuasion, not only suggests that the numerosity heuristic may affect attitude change, it also suggests that it is most likely to do so under conditions that preclude higher-order analysis (see also Hasher, Goldstein, & Toppino's, 1977, evidence that stimulus frequency may produce automatic inferences of referential validity). Thus, there is indirect evidence from a number of different sources suggesting the tendency to overinfer magnitude from multitude is a relatively effortless strategy.

On the other hand, except in cases in which people can instantly assess the numerosity of a relatively small number of objects without awareness (i.e., except in cases of subitizing; see Gallistel & Gelman, 1991), it is clear that counting is not completely effortless. Similarly, it seems highly unlikely that the operation of the numerosity heuristic fulfills all of the traditional criteria for automaticity (see Bargh, 1989; Logan, 1988, 1989). Future research should explore the effortlessness or automaticity of the numerosity heuristic in greater detail. The extent to which the operation of the numerosity heuristic is intentional and goal-dependent, for example, may have important implications for the degree to which people's after-the-fact (as opposed to on-line) judgments of quantity are influenced by numerosity (e.g., "Does Shirley have enough room at her house for three overnight guests?"). If the assessment of numerosity is every truly automatic, people may sometimes encode numerosity information without necessarily encoding other important sources of information. For some cases of memory-based judgments, this may mean that people cannot engage in systematic processing because they simply do not have access to the information needed to carry out systematic processing (e.g., "I think so; all I remember is that she lives in an eight room house").

Numerosity and Motivation

Although we can only speculate about the role of the numerosity heuristic in after the fact judgments, the results of the present studies suggest very clearly that when it comes to on-line judgments, depriving people of cognitive resources increases the likelihood that they will overinfer quantity from numerosity. Many researchers also assume that depriving people of sufficient motivation (e.g., by asking them to make judgments about uninvolving issues) increases the likelihood that they will take mental shortcuts, and there is some empirical support for this perspective (e.g., see Creyer, Bettman, & Johnson, 1990; Martin et al., 1990; Petty & Cacioppo, 1984; Tolman, Ritchie, & Kalish, 1946).

It may be instructive, however, to reconsider the role of motivation in systematic versus heuristic modes of judgment. Even those who are strongly motivated to make correct judgments may be forced to rely on "heuristic" decision rules if they lack either the expertise or the cognitive resources required to apply a demanding normative decision rule. In fact, in such cases, high levels of motivation may actually *increase* people's use of simplifying strategies.

In support of this perspective, Pelham and Neter (1993) have found that, when people are asked to make especially demanding judgments, high levels of motivation increase their reliance on simplifying heuristics (including the representativeness and the availability heuristics as well as the numerosity heuristic). For instance, in a study patterned directly after Experiment 1 of this paper, we manipulated participants' motivation to make accurate judgments by promising some participants monetary prizes for making especially accurate judgments. This study revealed that when systematic processing was especially difficult (i.e., in a difficult reassembly condition), greater levels of motivation facilitated less accurate (i.e., more heuristic) judgments. In contrast, when we made it easy for participants to engage in systematic processing (i.e., in an easy reassembly condition), greater motivation facilitated more accurate judgments (see also Chaiken, Liberman, & Eagly, 1989; Yerkes & Dodson, 1908). In short, the evidence presented in this report, in combination with other adaptive and contextual approaches to judgment, suggests that an appreciation of the adequacy of human judgment must be grounded in an understanding of the social, cognitive and motivational context in which people render their judgments.

Limitations on Numerosity

Although we have argued that numerosity plays a disproportionate role in a wide variety of judgments, it is clear that numerosity is only one of many cues for assessing quantity or probability. It is also clear that people rely heavily on the numerosity heuristic only when their cognitive resources are taxed in some important way. Moreover, we suspect that even people who are cognitively taxed will not always overinfer quantity from numerosity. For at least some judgments, people may possess intuitive thresholds for what constitutes a meaningful unit. The value of exceedingly small pieces of a whole that do not measure up to this threshold may either be trivialized or altogether ignored. Thus, naive bankers might fail to appreciate the value of a large jar of pennies, naive snackers might underestimate the number of calories they ingest by repeatedly nibbling on tiny portions of food, and naive smokers might fail to realize the cumulative damage of several years worth of individual cigarettes. If these things are true, we suspect that it is partly because people do not consider a mere penny, a lone peanut, or a single Pall Mall a meaningful psychological unit.

In the case of judgments of probability rather than quantity, the degree to which people make use of the numerosity heuristic may depend on the conjunctive versus disjunctive nature of the particular probability judgments people are asked to make. Bar-Hillel (1973) found that people typically overestimate the likelihood of compound *conjunctive* events (such as drawing a spade from a deck of cards on each of six independent trials) relative to equally probable simple events. However, she also found that people underestimate the likelihood of compound *disjunctive* events (such as drawing an ace on any of six independent trials) relative to equally probable simple events (see also Shaklee & Fischhoff, 1990). Although these findings are not necessarily incompatible with the existence of the numerosity heuristic,[5] they do make it clear that there is much more to the estimation of frequency or probability that simple numerosity. These caveats notwithstanding, the evidence presented in this report suggests that, more often than not, dividing a stimulus into multiple units will increase its subjective magnitude. Apparently, one of the first things people ask when they want to know "how much?" is "how many?".

CONCLUSION

About 5000 years ago, a group of Sumerian traders refined the *token-bullae* system into a pictographic accounting system designed to help them keep track of their economic transactions. Eventually, this revolutionary system developed not only into a formal number system but also into what some believe was the first written language (Green, 1981; Powell, 1981; Schmandt-Besserat, 1981). Whether these ambitious traders

[5] Recall that one likely explanation for the numerosity heuristic is the concave downward or S-shaped psychophysical function rule described earlier in this paper. This function rule predicts that people will typically overestimate small values and underestimate large values. Conjunctive compound events are typically very improbable, and people typically overestimate these probabilities. Disjunctive compound events are typically highly probable, and people typically underestimate these probabilities.

were formalizing an intuition shared by virtually all vertebrates or establishing a uniquely human tradition, it is clear that their accounting system has become woven indelibly into human thought. For precisely this reason, the numerosity of a stimulus may sometimes compel people to ignore other equally important cues for inferring quantity. At the same time, a potentially unique feature of human thought is the capacity to adjust our overly simplistic intuitions by engaging in effortful, analytical reasoning. As researchers continue to explore the parameters of human thought and judgment, it is hoped that they will examine both effortless and effortful forms of accounting and thereby help to delineate both the systematic biases and the systematic corrections that characterize human thinking.

REFERENCES

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology,* **2,** 1–9.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin,* **110,** 486–498.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology,* **41,** 258–290.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7, pp. 47–89). Orlando, FL: Academic Press.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance,* **9,** 396–406.

Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 3–51). New York: Guilford Press.

Ben Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica,* **47,** 89–104.

Birnbaum, M. H., & Veit, C. T. (1973). Judgmental illusion produced by contrast with expectancy. *Perception and Psychophysics,* **1,** 149–152.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes,* **14,** 3–17.

Capaldi, E. J., Miller, D. J., & Alptekin, S. (1989). Multiple-food-unit-incentive effect: Nonconservation of weight of food reward by rats. *Journal of Experimental Psychology: Animal Behavior Processes,* **15,** 75–80.

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (Vol. 5, pp. 3–40). Hillsdale, NJ: Erlbaum.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). New York: Guilford Press.

Creyer, E. H., Bettman, J. R., & Johnson, E. J. (1990). The impact of accuracy and effort feedback and goals on adaptive decision behavior. *Journal of Behavioral Decision Making,* **3,** 1–16.

Dunning, D., Griffin, D. W., Milojkovic, J., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology*, **58**, 568–581.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. London/New York: Cambridge University Press.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 552–564.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 330–344.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression-formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition* (2nd ed.). New York: McGraw–Hill.

Gallistel, C. R., & Gelman, R. (1991). Subitizing: The preverbal counting process. In W. Kessen, A. Ortony, & F. Craik (Eds.), *Essays in honor of George Mandler*. Hillsdale, NJ: Erlbaum.

Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology* (Vol. 3, pp. 167–230). New York: Wiley.

Gelman, R., Cohen, M., & Hartnett, P. (1989). To know mathematics is to go beyond thinking that "fractions aren't numbers." *Proceedings of the Eleventh Annual Meeting of the North American Chapter, International Group for Psychology of Mathematics Education*. New Brunswick, NJ.

Gelman, R., & Meck, E. (in press). Early principles aid early but not later conceptions of number. In J. Bideaud & C. Meljac (Eds.), *Les chemins du nombre*. Paris: Les Presses Universitaires De Lille.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189–211). New York: Guilford Press.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, **46**, 107–119.

Gilbert, D. T. (1993). The assent of man: Mental representation and the control of belief. In D. M. Wegner & J. W. Pennebaker (Eds.), *The Handbook of Mental Control* (pp. 57–87). Englewood Cliffs, NJ: Prentice-Hall.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, **54**, 723–740.

Ginossar, Z., & Trope, Y. (1987). Problem-solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, **52**, 464–474.

Ginsburg, N. (1978). Perceived numerosity, item arrangement, and expectancy. *The American Journal of Psychology*, **91**, 267–273.

Green, M. W. (1981). The construction and implementation of the cuneiform writing system. *Visible Language*, **XV**, 345–372.

Hartmann, G. W. (1935). *Gestalt Psychology: A survey of facts and principles*. New York: The Ronald Press Company.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, **16**, 107–112.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, **62**, 498–525.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492–527.

Logan, G. D. (1989). Automaticity and cognitive control. In J. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 52–74). New York: Guilford Press.

Maheswaran, D., & Chaiken, S. (1991). Promoting systematic processing in low motivation settings: Effect of incongruent information on processing and judgment. *Journal of Personality and Social Psychology*, **61**, 13–25.

Martin, L. L., Seta, J. J., & Crelia, R. A. (1990). Assimilation and contrast as a function of people's willingness and ability to expend effort in forming an impression. *Journal of Personality and Social Psychology*, **59**, 27–37.

McLaughlin, J. A. (1981). Development of children's ability to judge relative numerosity. *Journal of Experimental Child Psychology*, **31**, 103–114.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, **63**, 81–97.

Neisser, U. (1976). *Cognitive and reality*. San Francisco: Freeman.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, **28**, 28–38.

Newtson, D. (1980). An interactionist perspective on social knowing. *Personality and Social Psychology Bulletin*, **6**, 520–531.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, **7**, 44–64.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1990). The adaptive decision maker: Effort and accuracy in choice. In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 129–153). Chicago: University of Chicago Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, **43**, 87–131.

Pelham, B. W., & Neter, E. (1993). The harder I try, the behinder I get: The effects of motivation on judgement depend on the difficulty of the judgement. Under review.

Pelham, B. W., & Swann, W. B., Jr. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology*, **57**, 672–680.

Petty, R. E., & Cacioppo, J. T. (1984). The effect of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion. *Journal of Personality and Social Psychology*, **46**, 69–81.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration-likelihood model of persuasion. In L. Berkowitz (Ed.) *Advances in experimental social psychology* (Vol. 19, pp. 123–205). New York: Academic Press.

Poulton, E. C. (1982). Biases in quantitative judgements. *Applied Ergonomics*, **13**, 31–42.

Powell, M. A. (1981). Three problems in the history of cuneiform writing: Origins, direction of script, literacy. *Visible Language*, **XV**, 419–440.

Rothstein, H. G. (1986). The effects of time pressure on judgment in multiple cue probability learning. *Organizational Behavior and Human Decision Processes*, **37**, 83–92.

Russo, J. E. (1977). The value of unit price information. *Journal of Marketing Research*, **14**, 193–201.

Schmandt-Besserat, D. (1981). From tokens to tablets: A re-evaluation of the so-called "numerical tablets". *Visible Language*, **XV**, 321–344.

Shaklee, H., & Fischhoff, B. (1990). The psychology of contraceptive surprises: Cumulative risk and contraceptive effectiveness. *Journal of Applied Social Psychology*, **20**, 385–403.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and general theory. *Psychological Review*, **84**, 127–190.

Showers, C. (1992). Compartmentalization of positive and negative self-knowledge: Keeping bad apples out of the bunch. *Journal of Personality and Social Psychology*, **62**, 1036–1049.

Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.

Simon, H. A. (1981). *The sciences of the artificial*, 2nd ed. Cambridge, MA: MIT Press.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153–181.

Stich, S. P. (1990). Rationality. In D. N. Osheron & E. E. Smith (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, pp. 173–196). Cambridge, MA: MIT Press.

Swann, W. B., Jr., Hixon, J. G., Stein-Seroussi, A., & Gilbert, D. T. (1990). The fleeting gleam of praise: Cognitive processes underlying behavioral reactions to self-relevant feedback. *Journal of Personality and Social Psychology*, **59**, 17–26.

Taves, E. H. (1941). Two mechanisms for the perception of visual numerousness. *Archives of Psychology*, **265**, 1–48.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing Science*, **4**, 199–214.

Tolman, E. C., Ritchie, B. F., & Kalish, D. (1946). Studies in spatial learning; I. Orientation and the short-cut. *Journal of Experimental Psychology*, **36**, 13–24.

Traupmann, K. L. (1971). Acquisition and extinction of an instrumental running response with single- or multiple-pellet reward. *Psychonomic Science*, **22**, 61–63.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, **93**, 239–257.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**, 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

Wilder, D. A. (1977). Perception of groups, size of opposition, and social influence. *Journal of Experimental Social Psychology*, **13**, 253–268.

Wilder, D. A. (1978). Homogeneity of Jurors: The majority's influence depends on their perceived independence. *Law and Human Behavior*, **2**, 363–376.

Wolfe, J. B., & Kaplon, M. D. (1941). Effect of amount of reward and consummative activity on learning in chickens. *Journal of Comparative Psychology*, **31**, 353–361.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology of Psychology*, **18**, 459–482.

(Accepted April 25, 1993)