Available online at www.sciencedirect.com
ScienceDirect
Acta Psychologica 123 (2006) 337–354
acta psychologica
www.elsevier.com/locate/actpsy

Effects of processing fluency on comparative performance judgments

Rolf Reber [a,*], Beat Meier [b], Marie-Antoinette Ruch-Monachon [b], Mara Tiberini [b]

[a] *Department of Psychosocial Science, University of Bergen, Christiesgate 12, N-5015 Bergen, Norway*
[b] *Department of Psychology, University of Bern, Muesmattstr. 45, CH-3000 Bern 9, Switzerland*

Received 1 September 2005; received in revised form 1 February 2006; accepted 3 February 2006
Available online 22 March 2006

Abstract

Research has shown that performance predictions are biased by the impact of processing fluency. However, existing data are inconclusive with regard to comparative judgments of performance. In five experiments, participants in an easy condition gave more favorable comparative judgments than participants in a difficult condition. Participants judged their performance more favorably if they named colors of non-color words rather than non-matching color words (Experiment 1), if they had to generate six words of a category rather than 12 words (Experiment 2), if they had to run in place for 15 s rather than 2 min (Experiment 3), but the latter result holds only true if participants were not active in sports (Experiment 4). When 67% of the items in a recognition test were old words, participants thought that their recognition performance was better than when 33% of the items were old words, although recognition performance did not differ between groups (Experiment 5). We discuss this result in the light of recent theories about effects of processing fluency on judgments.

PsycINFO classification: 2340

Keywords: Processing fluency; Assessment; Performance; Cognitive task; Motor task

* Corresponding author. Tel.: +47 55 58 31 81; fax: +47 55 58 98 79.
E-mail address: rolf.reber@psysp.uib.no (R. Reber).

0001-6918/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2006.02.001

1. Introduction

Social comparison is ubiquitous; people compare themselves to others on attractiveness, wealth, health, or performance, to name just a few examples. The outcome of comparison processes have been shown to depend both on social goals, such as accurate self-evaluation or self-enhancement (see Wood, 1989), and cognitive factors, such as differential attention or anchoring and adjustment (Chambers & Windschitl, 2004). Chambers and Windschitl discussed also the impact of accessibility, which often is phenomenally experienced as ease of recall (see Schwarz, 1998). We shall examine how ease of performance influences comparative performance evaluations.

For example, it is easier to run 15 s than to run 2 min. Is it possible that this ease of running influences judgments of how good one runs in comparison to others? We shall examine this and related questions in this article. The key term is processing fluency, which is the subjective ease with which information is processed (see Reber, Wurtz, & Zimmermann, 2004). Instances of processing fluency are perceptual fluency, the subjective ease of processing sensory input, or retrieval fluency, the subjective ease of recalling information from memory. We focus on one aspect of processing fluency: On the subjective ease with which a task can be performed. As Reber, Wurtz, et al. (2004) have shown in perceptual tasks, the objective speed at different stages contributed jointly to the subjective experience of fluency. The subjective ease of processing experienced during a task is a function of a number of factors, such as amount of information retrieved from memory, expressed as number of letters in a feeling of knowing task (Koriat, 1993) or as numbers of retrieved items (Williams & Durso, 1986), performance per unit of time (Morris, 1990), or objective task difficulty (see Schwarz, 1998). In five experiments, we demonstrated that effects of processing fluency cannot be reduced to effects of amount of information, performance per unit of time, or objective task difficulty: Subjective experience of ease resulted in more positive comparative judgments of performance even when amount of information was lower in the easy than in the difficult condition, and when performance per time unit and task performance were constant, suggesting that our manipulations of processing fluency did not merely reflect these other variables.

We first discuss studies that have demonstrated the impact of retrieval fluency on different metacognitive judgments, such as feeling of knowing and judgments of learning, before we show that these studies are inconclusive when it comes to the influence of processing fluency on comparative performance judgments.

1.1. Effects of retrieval fluency on metacognitive judgments

Retrieval fluency—the subjective ease with which people can retrieve information from memory—is related to metacognitive judgments, such as feeling of knowing (e.g., Koriat, 1993) and judgments of learning (e.g., Hertzog, Dunlosky, Robinson, & Kidder, 2003; Matvey, Dunlosky, & Guttentag, 2001; Rawson & Dunlosky, 2002; Simon & Bjork, 2001, 2002). Feeling of knowing is the phenomenal experience that the information to be remembered is stored in memory even if one is currently unable to retrieve this information (Koriat, 1993). Judgments of learning are predictions of retrieval after the encoding of the relevant material. Rawson and Dunlosky (2002) have found that ease of processing, manipulated by text coherence, influenced predictions of text recall. Hertzog et al. (2003) demonstrated effects of ease of processing at encoding on judgments of learning in a pair associate learning task. Benjamin, Bjork, and Schwartz (1998) exploited the phenomenon of experienced retrieval fluency contradicting memory performance: Retrieval fluency is high, but memory performance low; or retrieval fluency is low, but memory performance high. For example, in answering general knowledge questions (Gardiner, Craik, & Bleasdale, 1973) people experience higher retrieval fluency if they can easily retrieve the answer to a question than if they have difficulties to find the answer. However, people are less likely to recall the easy answers than those they had to think about before finding them. Participants in Benjamin et al.'s study had to answer general knowledge questions and then to predict how likely they were to reproduce given answers in a test that followed 20 min later. As predicted by a retrieval fluency account, participants thought that they would do best when the answer was found easily rather than with difficulty; actual retrieval, however, showed the inverse pattern, as already demonstrated by Gardiner and his colleagues. Simon and Bjork (2001, 2002) found related dissociations of predictions and performance in motor tasks.

Winkielman, Schwarz, and Belli (1998) instructed their participants to recall childhood memories. One group had to recall four such memories, another group 12. After recall, the participants were asked "Regarding your childhood memory, are there large parts of your childhood after age 5 which you can't remember". As predicted by an ease of recall hypothesis, those who recalled four instances were less likely to think that large parts of their childhood memory were lacking than those who recalled 12 instances. Winkielman et al. concluded that participants used retrieval fluency as information to arrive at their conclusion.

All these studies examined predictions or assessments for the participants themselves, demonstrating an effect of retrieval fluency on self-judgments of learning or of one's own childhood memories. Kelley and Jacoby (1996) examined a situation where participants, after having solved a task, had to judge how difficult the task was for others. They asked their participants how difficult anagrams were for other people to solve. Some solution words of the anagrams were shown in a first session. Later, participants had to solve the given anagrams. Familiar anagrams whose solution words had been shown before were solved faster than new anagrams and judged as being easier to solve for other people. The authors concluded that the participants attributed higher fluency in solving familiar anagrams to the ease with which others are able to solve these anagrams. In line with this conclusion, groups did not show this judgmental bias either when they were tested with new anagrams only, or when they were first given the solution and then the anagram in order to judge the difficulty for other people, depriving them of the phenomenal experience of ease of solving anagrams.

These studies show that processing fluency influences both self- and other-related judgments in regard to memory performance. The question arises how processing fluency influences comparative performance judgments.

1.2. Comparative social judgments

All studies that demonstrate effects of processing fluency on judgments of learning (Benjamin et al., 1998; Hertzog et al., 2003; Rawson & Dunlosky, 2002; Simon & Bjork, 2001, 2002) or inferences about task difficulty (Kelley & Jacoby, 1996) or the quality of childhood memories (Winkielman et al., 1998) employed judgments that did not include a comparison of participants with peers. An example of a comparative performance judgment is when students do a recognition test and then evaluate how they performed in comparison to representative other students. How would differences in experienced processing fluency influence comparative judgments? We consider two possible outcomes.

First, if people give performance-related judgments on the basis of retrieval fluency, they may reason that a task is easy or difficult either for all participants alike or for them as an individual. For example, if participants in the Benjamin et al. (1998) study erroneously concluded that they would be able to retrieve an answer to a general knowledge question in the future because they were able to easily retrieve it in the first place, they might conclude that others would also retrieve answers that were easy to find. Therefore, manipulating retrieval fluency may affect self-related and other-related judgments of learning alike, but not comparative judgments: As both self-related judgments (Benjamin et al., 1998; Hertzog et al., 2003; Rawson & Dunlosky, 2002; Winkielman et al., 1998) and other-related judgments (Kelley & Jacoby, 1996) change in lockstep with increasing retrieval fluency, comparative judgments may stay constant. This would result in a null effect of processing fluency. If so, manipulations of performance fluency would not have any effect on comparative performance evaluations.

Alternatively, people may show an egocentric bias in comparative judgments by anchoring their judgment on their experienced processing fluency, but not being able to adjust it accordingly (see Chambers & Windschitl, 2004, for a discussion of potential mechanisms underlying comparative judgments). The increase in evaluation with increasing processing fluency may be greater for self-related judgments than for other-related judgments, resulting in higher comparative performance judgments with increasing processing fluency. Previous studies do not tell us what would happen in the case of comparative performance judgments. Two studies become especially relevant here: Matvey et al. (2001) presented findings in support of the notion that Judgments of Learning by observers are partly mediated by analytic processes. If people in comparative judgments use both analytic and non-analytic sources—theories and subjective experiences—for judging their own performance, but mainly analytical sources to judge the performance of others, they may show a steeper increase for the self-related judgment than for the other-related judgments, resulting in an increase in the comparative judgment.

Kruger (1999, Experiment 2) gave his participants a bogus test of "integrative orientation". In this test, adapted from Mednick and Mednick (1967), participants were given word triads and asked to provide a solution word that was related to all three words. One group got easy triads; the other group got difficult triads. The group given 10 easy triads solved 8.1 on average; the group given 10 difficult triads solved 1.9 on average. When asked about their ability to solve such triads, the group given the easy triads gave an average percentile estimate of 63.4, which was significantly above the 50% mark; the group given the difficult triads gave an average percentile estimate of 41.9, which was significantly below 50%. Kruger explained this finding with egocentrism in judgments that result in incomplete adjustment after anchoring to the remembered performance.

Why should participants be egocentric in judging their ability? One reason may be that to participants, information about themselves is more available than information about others (e.g., Kruger, 1999; see Reber, 2004). Indeed, in Kruger's study, the participants' comparative judgments were predicted better by the judgment of their own ability than by the judgment of the ability of others. If comparative judgments in our study were egocentric, that is, if they were more influenced by self-related than by other-related performance evaluations, we would find an effect of our processing fluency manipulation on these judgments.

Kruger's (1999) studies were not aimed at testing the assumption that processing fluency influences judgments. However, in his Experiment 2, he demonstrated that participants who got the easy test gave above-average judgments of ability whereas participants who got the difficult test judged themselves to be less able than the average in performing this kind of task. This finding is open to various interpretations: Kruger assumed that participants might have anchored their judgment on their own skill in solving the task so that they arrived at higher scores after insufficient adjustment if they inferred high skill from having solved many tasks, as in the easy condition, than if they were in the difficult condition and concluded low skill from having solved only few tasks. Alternatively, Kruger's findings could be interpreted as an effect of processing fluency: Students in the easy test solved the triads more fluently than participants in the difficult test, and then used processing fluency as a basis for their comparative ability judgment. Therefore, even a processing fluency explanation of this study leaves an ambiguity between ease of processing and amount of processed information (see Schwarz, 1998; Schwarz et al., 1991).

In sum, we are left with findings that are inconclusive for comparative performance judgments: Research on self-related and other-related metacognitive judgments leave open the question how processing fluency affects comparative performance judgments. The study by Kruger (1999) is open to various interpretations; if construed in terms of processing fluency, it does not address the issue whether participants used ease or amount of processed information.

To fill this gap we conducted five experiments to test the idea that ease of processing contributes to comparative performance judgments. There are two kinds of comparative judgments: Direct and indirect ones (see Chambers & Windschitl, 2004). Direct comparative judgments ask one question: Participants have to judge themselves on some dimension, such as performance, compared to average others. Indirect comparative judgments ask two questions: Participants have to judge first their own performance, and second the performance of average others of a representative group. In all experiments, we chose to assess comparative judgments with the direct method, for two reasons: First, the direct method is simpler than the indirect method in that we can directly analyze the score of one question and do not have to calculate a difference score. As the influence of processing fluency on both self-related and other-related judgments has already been demonstrated, we are interested in a simple measure of comparative judgments. Second, the direct method yields stronger effects than the indirect method (see Chambers & Windschitl, 2004). As we aimed at documenting an effect if it exists, we liked to avoid committing a type II error. In addition to the performance measure, which served as a proxy for objective task difficulty, ease of processing was measured as judged difficulty and comparative performance judgments required participants to evaluate their performance in comparison to imagined fellow students.

In Experiment 1, participants were given a color naming task. One group had to name the colors of neutral words (non-color words), another group to name the colors of non-matching color words, as in the Stroop task (Stroop, 1935; see MacLeod, 1991 for an overview). Naming the color of neutral words is easier than naming the color of non-matching color words. From the analysis given above, two predictions are possible: First, if our manipulation of performance ease affected evaluations of one-self and of others in the same way, we would predict a null effect. Alternatively, if participants' self-evaluations were affected more by our manipulation of processing fluency than evaluations of the others, participants who performed the easy task would be expected to give higher comparative performance evaluations than participants who performed the difficult task. Our study is a conceptual replication of Kruger (1999, Experiment 2) in which an easy task led to higher comparative performance evaluations than a difficult task. However, in our Experiment 1, participants did not get any information of what constitutes an average performance. Therefore, there was no external standard to anchor comparative performance evaluations, and an effect of our manipulation would suggest an effect of processing fluency and excludes the intentional use of external standards. In Experiment 2, participants were given a word generation task. In contrast to Experiment 1, both groups had to solve exactly the same type of task, the only difference being that in the difficult condition, participants had to generate more words than in the easy condition. In this experiment, we disentangled ease of processing from amount of processed information by comparing groups that had to generate 6 or 12 stimuli, respectively (see Schwarz, 1998). Generating six items is easier, but fewer in number than generating 12 items. If processing fluency influences comparative performance judgments, we expected that those who had to generate six words would evaluate their own performance more favorably than those who had to generate 12 words. If, however, the amount of processed information is the decisive variable, then the opposite pattern could be expected. In Experiment 3, we examined how processing fluency during the execution of a motor task affects performance evaluations. Participants who were not active in sports had to run for either 15 s or 2 min and were then asked to rate their performance. In this experiment, we held running speed constant, thus excluding the possibility that differences in objective performance resulted in a difference of rated performance. We expected again—if processing fluency influenced comparative judgments—that the group with the easy task would give more favorable evaluations of their performance than the group with the difficult task. Experiment 4 used an identical design, but in addition, we manipulated involvement in sports activities in order to examine the idea that such involvement results in the use of amount of running as information. Finally, in Experiment 5, we manipulated processing fluency by varying the proportion of old words in a recognition test. There is compelling evidence that previous exposure to words increases perceptual fluency when these words are shown again (e.g., Jacoby & Dallas, 1981; Whittlesea, 1993; Whittlesea & Williams, 1998, 2000). We expected that experienced fluency during a recognition test is higher and thus—if processing fluency influences self-related and other-related judgments differently—yield more favorable comparative judgments if the proportion of old items is .67 rather than .33.

In each experiment, we also assessed difficulty and well-being. Judged difficulty was used as a measure of processing fluency. Well-being was assessed to test whether the easy group feels better and whether this positive feeling translates into more favorable performance evaluations as both fluency and positive moods are signals of positively marked feelings (see Reber, Schwarz, & Winkielman, 2004; Schwarz, 2002; Winkielman, Schwarz, Fazendeiro, & Reber, 2003). All experiments used identical scales for the assessment of performance, judged difficulty, and well-being.

2. Experiment 1

2.1. Method

2.1.1. Participants

Forty-three undergraduate students (5 men and 38 women) from a Swiss University participated in the experiment for partial course credit. Their age was between 19 and 41 years ($M = 23.7$). Twenty-two participants were randomly assigned to the easy condition, 21 participants to the difficult condition. Participants were tested individually in a psychology laboratory of the university.

2.1.2. Materials and procedure

Participants first had to name colors of rectangles. This task was used as a baseline condition. The next task was the Stroop task: Participants in the easy condition had to name one by one the colors of 36 neutral words that were arranged in a 4×9 matrix. Participants in the difficult condition had to name the colors of 36 color words with the same arrangement. Color and meaning of the words never matched. The experimenter recorded time and number of errors for both the baseline task and the experimental task.

After the Stroop task, participants had to estimate their performance in this task in comparison to other participants by putting a mark on an unnumbered scale. They were asked: " In comparison to other participants, my performance in this task is…" The scale ranged from "very bad" at the lower end through "average" in the centre to "very good" at the upper end. The scale was one hundred millimeters wide so that we were able to measure where the participants put a mark and calculate the percentile as a measure of their evaluation. If the mark was 57 mm from the lower end of the scale, this meant that the participant judged himself or herself to be at least as good as 57% of the other participants, or—in other words—that the performance of 57% of the other participants was supposed to be worse than his or her performance had been. After this comparative judgment, participants had to indicate how much time they needed for the task and then to rate difficulty and wellbeing on a scale from 1 to 9.

2.2. Results and discussion

2.2.1. Baseline task

Mean time for naming colors of rectangles was $M = 18.0$ (SD = 3.6) seconds for the easy group and $M = 18.5$ (SD = 3.7) seconds for the difficult group. This difference was not significant, $t(41) = .47$, $p = .644$. Neither differed the number of errors for the two groups, $M = .27$ (SD = .55) and $M = .38$ (SD = .50), $t(41) = .68$, $p = .503$. As there was no group difference in the baseline task, we analyzed the performance data of the subsequent Stroop task directly, without computing the differences to the baseline conditions.

2.2.2. Performance and performance evaluation

Mean time to perform the task was $M = 21.8$ (SD = 4.9) seconds for the easy group and $M = 30.7$ (SD = 4.9) seconds in the difficult group. Number of errors was $M = .50$ (SD = .91) and $M = 1.52$ (SD = 1.60), respectively. These differences were significant, $t(41) = 5.90$, $p < .001$, and $t(41) = 2.59$, $p = .013$, respectively. As expected, participants in the easy group were faster and made fewer errors than participants of the difficult group. Participants in the easy group estimated that they used $M = 31.6$ (SD = 12.9) seconds, participants in the difficult group gave an estimate of $M = 48.0$ (SD = 29.5). Participants in both groups overestimated the time they needed to name the colors, as indicated by the difference between estimated time and objective time, $t(21) = 3.54$, $p = .002$, and $t(20) = 2.66$, $p = .015$, respectively, however, there was no group difference in overestimation of the time needed, Ms = 9.77 and 17.33, $t(41) = 1.09$; the median for overestimation of time by the two groups was 9.0 and 5.0, respectively, suggesting that the distribution of

Table 1
Means (and standard deviations) for performance evaluation and judged difficulty for all five experiments

Experiment	Performance evaluation (%)		Judged difficulty	
	Easy	Difficult	Easy	Difficult
1	60.7 (13.3)	48.0 (19.8)	4.18 (1.53)	6.52 (1.12)
2	47.0 (18.7)	32.3 (15.8)	4.23 (1.38)	4.77 (1.26)
3	63.0 (19.0)	48.3 (19.3)	1.70 (.92)	3.05 (1.67)
4 Inactive	62.5 (14.1)	49.5 (14.3)	1.95 (.76)	2.80 (1.94)
4 Active	64.0 (14.7)	71.5 (13.9)	2.05 (1.32)	2.15 (1.42)
5	67.4 (12.5)	58.9 (14.1)	3.85 (1.38)	4.61 (1.33)

the difficult group was skewed. Therefore, biases in performance judgments were not caused by biases in time estimation.

Means and standard deviations for performance evaluation and judged difficulty are shown in the first row of Table 1. Both differences were significant, $t(41) = 2.48$, $p = .017$ and $t(41) = 5.69$, $p < .001$. Participants in the difficult group gave significantly higher difficulty ratings than participants of the easy group, suggesting that our ease of processing manipulation was successful. The difference in well-being was not significant, $M = 6.9$ (SD = 1.2) for the easy group and $M = 6.6$ (SD = 1.5) for the difficult group, $t(41) = .82$.

Our findings suggest that participants used processing fluency as information for evaluating their own performance. One could argue that the group difference in performance evaluation emerged from the different tasks the participants were given. Although there is no obvious theoretical reason why naming the colors of neutral words should result in more favorable evaluations of one's own performance than naming the colors of color words, it is not possible to disentangle the contribution of the two factors—ease of task and type of task—to performance evaluations. The next four experiments were designed to examine whether the contribution of ease of processing to evaluations could be reduced to factors often associated with processing fluency, such as amount of information, performance per unit of time, and task difficulty. We therefore used the identical task for both groups. In the next three experiments, we relied on a technique Schwarz et al. have used in order to manipulate ease of processing (e.g., Schwarz et al., 1991; see Reber, 2004 & Schwarz, 1998 for overviews). In one of their studies, they instructed one group of participants to list six instances of past behavior where the participants were self-assertive; another group of participants was instructed to list 12 instances of self-assertive behavior. After recall of the instances, participants were asked how self-assertive they were. Participants who had to list six behaviors judged themselves to be more self-assertive than participants who had to list 12 self-assertive behaviors. Schwarz et al. (1991) concluded that their manipulation of ease of recall affected self-judgments of assertiveness. In our studies, ease of task was manipulated by the amount of items participants had to generate (Experiment 2) or by the time participants were given to execute the task until they were interrupted (Experiments 3 and 4); amount of information in Experiment 2 and time in Experiments 3 and 4 were inversely related to ease of performance.

3. Experiment 2

We used a word generation task that is widely used in neuropsychology and known to become more difficult with time (see Spreen & Strauss, 1998). We assumed that participants did not have too much knowledge about how they themselves or others perform in this task and hence were lacking a standard of comparison. Therefore, participants might use processing fluency as information when they have to evaluate their performance. If so, participants who had to generate six words were expected to provide more favorable evaluations of their performance than participants who had to generate 12 words. If, however, participants used number of retrieved words as information for their performance, the group that had to generate 12 words was predicted to give higher comparative judgments of performance than the group that had to generate six words.

3.1. Method

3.1.1. Participants

Sixty-one students from a Swiss community college (30 men and 31 women) participated as volunteers in the experiment. Their age was between 18 and 21 years ($M = 19.0$). Thirty participants were randomly assigned to the group that had to generate six words, 31 participants to the group that had to generate 12 words. Participants were tested individually in a small room of the community college.

3.1.2. Materials and procedure

Participants were instructed to generate as fast as possible words that began with a certain letter. About one third of the participants got the letter B, another third the letter S, and the remaining participants the letter T. Oral production of the words was recorded on a tape-recorder. Participants were not told that they would be stopped after 6 or 12 words, respectively. The experimenter silently counted the number of generated words and stopped the participants after 6 or 12 words (according to experimental condition), and measured the time.

After having been stopped, the participants had to estimate their performance in this task in comparison to other participants on an unnumbered scale, identical in design to the one used in Experiment 1. Moreover, we assessed subjective estimates of number of words and time, and subjective difficulty of the task and well-being.

3.2. Results and discussion

Number of words was fixed at 6 words or 12 words, respectively. The respective estimated number of words were $M = 6.3$ (SD = 1.3) and $M = 12.4$ (SD = 4.2), which was not significantly different from the objective measures, $t(29) = 1.14$ and $t(30) = .52$. Mean time for word generation was $M = 25.0$ (SD = 13.4) seconds for the group that generated six words and $M = 83.9$ (SD = 45.2) seconds for the group that generated 12 words. The respective estimated times were $M = 25.4$ (SD = 18.9) seconds and $M = 73.8$ (SD = 45.9), which was not significantly different from the objective measures, $t(29) = .17$ and $t(30) = 1.53$.

Means and standard deviations for performance evaluation and judged difficulty are shown in the second row of Table 1. As in Experiment 1, participants with the easier task gave higher evaluations of their own performance than participants with the more difficult task, $t(59) = 3.32, p = .002$. This group difference replicates the finding of Experiment 1 even though the type of task was the same for both groups. There were no biased estimates of time used or number of words generated. Therefore, biases in comparative performance evaluations were not caused by biased time or frequency estimates. This time, the responses to the difficulty question were not significantly different, $t(59) = 1.60$, although the means indicated that the difficult group found the task more difficult than the easy group. Again, well-being did not show any group differences, $M = 6.6$ (SD = 1.4) for the easy group and $M = 6.0$ (SD = 1.8) for the difficult group, $t(59) = 1.46$.

This experiment provides further support for the notion that processing fluency influences comparative performance judgments, and that the effect is not due to amount of processing, one indicator of processing fluency used in earlier research (e.g., Koriat, 1993; Williams & Durso, 1986). It is possible, however, that people base their judgment on an assessment of the amount of words they were able to generate per time unit (see Morris, 1990, for a related measure). If so, lower performance evaluations may be due to the fact that the difficult group generated fewer words per time unit, and not due to differences in processing fluency. Experiments 3 and 4 were designed to circumvent this possibility by holding performance constant.

4. Experiment 3

This experiment aimed at creating constant performance conditions. As outlined in the discussion of Experiment 2, it is possible that participants in the difficult condition may have evaluated their performance less favorably because they performed less well in terms of amount of generated words per time unit. In this experiment, participants had to run for 15 s or for 2 min. Performance (defined as number of steps per time unit) was held constant for both groups by using a metronome to provide the pace of running. Participants had to perform 168 steps per minute, resulting in 42 steps totally for those participants who ran 15 s and 336 steps for those running 2 min. If we find a group difference, we have further evidence that processing fluency and not difference in performance affects the evaluation. Moreover, by letting participants run, we generalize the findings from cognitive tasks (Experiments 1 and 2) to the domain of motor tasks.

4.1. Methods

4.1.1. Participants

Forty undergraduate students from a Swiss University (10 men and 30 women) participated in the experiment for partial course credit. Their age was between 19 and 44 years ($M = 23.6$) and they were tested individually in a psychology laboratory of the university. Half of them were in the group that ran for 15 s, the other half in the group that ran for 2 min.

They were pre-selected on the basis of their sportive activity. Only students who indicated an activity of less than two hours per week qualified for participation in the study. This was done because fitness centers and sports clubs provide fitness assessments. Some of these assessments express the outcome in terms of percentiles within a representative sex and age group, a measure similar to the one used in this experiment. There is evidence that people do not use processing fluency as information if they can retrieve a judgment directly (e.g., Haddock, Rothman, Reber, & Schwarz, 1999). Therefore, people who know how fit they are may directly retrieve their fitness assessment outcomes and base their comparative performance rating on it without resorting to processing fluency.

Another criterion was health. Before the experiment started, students had to complete a form that asked them whether there were any medical conditions that prevented them from running. The following medical conditions were mentioned explicitly on the form: Heart problems, asthma, problems with joints, and pregnancy. Only those students who responded that they could run and answered "no" to all specific questions were qualified to participate in the experiment; the others were given credit, thanked, debriefed, and dismissed.

4.1.2. Materials and procedure

Participants were instructed to take off their shoes and to run in place until the experimenter stopped them. In order to keep performance constant, the pace was set by a metronome that was tuned to 168 beats per minute, resulting in 42 steps for participants who ran for 15 s and in 336 steps for participants who ran for 2 min. Participants had to keep this pace until they were stopped; none of the participants had any problem with maintaining this steady pace.

After being stopped, the participants had to estimate their performance in this task in comparison to other participants on an unnumbered scale, the same as the one described in the former experiments. Then, we assessed subjective estimates of time and number of steps, and subjective task difficulty and well-being.

4.2. Results and discussion

Performance in this experiment was held constant by the pace of the metronome. Means and standard deviations for performance evaluation and difficulty are shown in the third row of Table 1. Participants with the easier task gave higher evaluations of their own performance, but lower difficulty ratings, than participants with the more difficult task $t(38) = 1.42, p = .021$, and $t(38) = 3.16, p = .003$.

Again, biased estimates of running time and number of steps did not explain biased performance judgments: The easy group, which overestimated their performance, actually underestimated both time and number of steps, $M\text{s} = 32.1$ (SD = 16.4) and 65.8 (SD = 26.7), $t\text{s}(19) = 4.67$ and 3.99, $p\text{s} < .001$. The difficult group, which judged to be about on average level, showed a non-significant tendency to overestimate both time and number of steps, $M\text{s} = 130.5$ (SD = 58.7) and 271.5 (SD = 210.1), $t\text{s}(19) = .80$ and 1.37. Average well-being was identical in both groups, $M = 7.2$ (SD = 1.1) for the easy group and $M = 7.2$ (SD = 1.5) for the difficult group, $t(38) = .00$.

Although performance—steps per time unit—was the same for both groups, the experimental manipulation influenced performance evaluations. This finding further bolsters the assumption that processing fluency plays a crucial role in the process of judging comparative performance, and that this effect cannot be reduced to an effect of performance per time unit. Critics may argue, however, that the task is less difficult and demanding for participants running for 15 s than for participants running for 2 min, and that the participants' judgments were based on task difficulty rather than experienced processing fluency.

In Experiment 4, we added a group active in sports and compared their judgments to the judgments of participants who were inactive in sports. If the difficulty of the task influenced comparative performance judgments, we would expect that both activity groups judge themselves to be worse when running for 2 min than when running for 15 s. However, there is good reason to predict that participants active in sports show a different pattern.

5. Experiment 4

Experiment 3 examined people who were not active in sports and showed that these participants used ease of processing as information for their judgment. Schwarz et al. have shown, however, that people may use amount of processing as information when they are highly motivated to process information (see Schwarz, 1998). Rothman and Schwarz (1998), for example, demonstrated that participants without a history of heart disease used retrieval fluency when assessing their risk for heart disease, whereas participants with a history of heart disease used amount of retrieved information for the same task. In a similar vein, people inactive in sports may use ease of processing, as demonstrated in Experiment 3, but people active in sports may use amount of information because they presumably are highly motivated to process information relevant to their sport performance.

Alternatively, participants active in sports may retrieve their fitness level directly, as discussed above; in this case, we would expect no effect of processing fluency on performance judgments. In sum, we predict that participants inactive in sports are influenced by ease of processing, whereas participants active in sports either retrieve their performance level directly (see Haddock et al., 1999), or they are motivated to systematically process information related to their sport performance, resulting in the use of amount of information for the comparative performance judgment. If participants simply anchored their self-assessment on task difficulty, we would expect that participants in both sports activity conditions score lower on comparative performance judgments when they run longer.

5.1. Methods

Eighty undergraduate students from a Swiss University participated in the experiment for partial course credit. Their age had to be between 18 and 30 years and they were tested individually in a psychology laboratory of the university. Forty participants were inactive in sports, according to the same criteria as in Experiment 3. Forty were active in sports: They trained at least twice a week in a sport club and participated regularly in competitions. Half of the participants in each group were in the condition that ran for 15 s, the other half in the condition that ran for 2 min.

Materials and procedures were the same as in Experiment 3, with one exception: We did not assess subjective time and number of steps, because they did not contribute to biased performance judgments in Experiment 3.

5.2. Results and discussion

Performance in this experiment was held constant by the pace of the metronome. Means and standard deviations for performance evaluation and difficulty are shown in the fourth and fifth row of Table 1. We performed a two-way analysis of variance, with the factors sportive activity and running time manipulated between subjects. For performance evaluation, there was a significant main effect of sportive activity, $F(1,76) = 13.62, p < .001$;

participants active in sports scored higher than participants inactive in sports, $M = 67.8$ (SD = 14.6) and $M = 56.0$ (SD = 15.5). This main effect was qualified by a significant sportive activity × running time interaction, $F(1,76) = 10.37$, $p = .002$. Participants inactive in sports gave higher comparative performance judgments if they had to run for 15 s rather than for 2 min, $t(38) = 2.89$, $p = .06$. In contrast, participants active in sports provided higher comparative performance judgments if they had to run for 2 min rather than 15 s, but this difference failed to be significant, $t(38) = 1.66$, $p = .11$.

The same two-way analyses of variance were run for difficulty and well-being as dependent variables. Difficulty did not show any significant effects; we found a significant effect of running time on difficulty in Experiment 3, where we tested only participants inactive in sports. In the present experiment, the same effect for the inactive group was marginally significant, $t(38) = 1.83$, $p = .075$. Well-being did only show a marginally significant effect of activity, $F(1,76) = 3.29$, $p = .074$; active participants felt somewhat better than inactive ones, $M = 7.85$ (SD = 1.03) and $M = 7.38$ (SD = 1.29), respectively. The other effects were not significant, $F(1,76) < 1.55$.

The results show that participants inactive in sports use ease of running as information for their comparative performance judgment. It is unclear whether participants active in sports used amount of running as information for the same judgment, or relied on information that they recalled directly. The direction of the difference is towards the use of amount of information, but failed to be significant by a relatively narrow margin, precluding definitive conclusions for those participants active in sports. Nonetheless, the experiment definitely revealed that participants did not simply anchor their judgment on task difficulty. Interestingly, well-being could not explain the interaction found in this experiment: Well-being was marginally higher for participants who were active in sports than those who were not, but it was not affected by the running time condition.

6. Experiment 5

Instead of artificially creating constant performance conditions, we designed an experiment in which we expected differences in processing fluency, but not in performance. The latter measure served as a proxy for task difficulty: We expected that the more difficult task would result in lower recognition performance. Specifically, participants were given a list of words that they had to judge for affective valence. They were then given a recognition test; in the easy group, 80 of 120 words were old words; in the difficult group, 40 of 120 were old words. There is a plethora of evidence that old items are easier to process than new items (e.g., Jacoby & Dallas, 1981; Whittlesea, 1993; Whittlesea & Williams, 1998, 2000). Therefore, participants in the easy group, who got a higher proportion of old words, were expected to experience higher processing fluency than participants in the difficult group. However, we did not expect that participants in the two groups would differ in their recognition test performance, that is, that the recognition task has the same difficulty for both groups.

6.1. Methods

6.1.1. Participants

Ninety-two undergraduate students (14 men and 78 women) from a Swiss University participated in the experiment for partial course credit. Their age was between 19 and 42 years ($M = 22.8$) and they were tested individually in a psychology laboratory of the university. Forty-six participants were assigned to the easy group which received 67% old items in the recognition test and 46 participants were assigned to the difficult group which received 33% old items in the recognition test. Two female participants of the difficult group had to be excluded from the analysis because their recognition performance in terms of the discrimination Index *Pr*, derived from the two-high threshold model of recognition (Snodgrass & Corwin, 1988), was around 0 and differed by more than 2.5 standard deviations from the mean performance.

6.1.2. Materials and procedure

A list of 90 words was presented on paper, and participants had to rate the affective valence of each word. After this incidental learning task, participants were presented 90 words, either 60 old and 30 new words (easy group) or 30 old and 60 new words (difficult group). They had to indicate for each word whether it had been shown before or not.

After the recognition test, participants had to estimate their performance in this task in comparison to other participants, as in the previous experiments. They then judged the difficulty of the task and responded to the well-being question.

6.2. Results and discussion

Means and standard deviations for the performance evaluation and judged difficulty are shown in the last row of Table 1. The groups did not differ in recognition performance, neither in terms of hits, $Ms = .90$ (SD = .06) versus .89 (SD = .07), $t(88) = .85$, false alarms, $Ms = .14$ (SD = .09) versus .12 (SD = .09), $t(88) = .94$, nor the discrimination index *Pr*, $Ms = .76$ (SD = .11) versus .77 (SD = .09), $t(88) = .26$. As in the former experiments, participants with the easier task (67% old words), gave higher evaluations of their own performance than participants with the more difficult task (33% old words), $t(88) = 2.98$, $p = .003$. Again, groups differed in difficulty judgments, $t(88) = 2.67$, $p = .009$, but not in well-being, $M = 6.86$ (SD = 1.31) for the easy group and $M = 7.04$ (SD = 1.01) for the difficult group, $t(88) = .73$.

Although recognition performance was comparable for both groups, the experimental manipulation only influenced performance evaluations. Interestingly, subjective difficulty judgments differed between groups although task difficulty in terms of recognition performance did not differ. This finding further bolsters the assumption that processing fluency plays a crucial role for performance judgments.

7. General discussion

In this study we investigated the influence of processing fluency on comparative performance evaluations. Processing fluency was manipulated by contrasting an easy and a difficult task condition. Comparative performance evaluations were assessed with the direct method, by asking participants to evaluate their performance in comparison to other participants of the experiment. In five experiments we consistently showed that processing fluency affected performance evaluations with higher comparative judgments under higher fluency conditions. In all experiments, we took care that the conditions were as similar as possible for the two groups, except for the critical variable in order to manipulate processing fluency. In each experiment, we excluded potential sources of confounds.

In Experiment 1, participants had no external standard to anchor their judgment on. In Experiment 2, we removed a possible confound of processing fluency and type of task by disentangling ease of generating words from number of words generated, a technique that has proved useful in research in social psychology (see Schwarz, 1998). However, it can still be possible that people assessed words per time, resulting in judgments that were not based on processing fluency. In order to remove this potential influence, we held performance constant in Experiment 3, and again the predicted higher performance evaluation for easy than for difficult conditions materialized. However, as running in place is a physically involving activity, one might argue that inexperienced participants judged their performance due to the difficulty of the task, not due to experienced fluency during running. In Experiment 4, we grouped participants according to sportive activity. This experiment replicated the findings from Experiment 3 for those participants not active in sports. In contrast, participants who were active in sports scored marginally higher on comparative performance evaluations when they ran for 2 min than when they ran for 15 s, excluding the explanation that participants use objective task difficulty as information for their judgment. In Experiment 5, we introduced a more subtle and physically less involving manipulation, so that task difficulty for both groups was expected to be the same. Participants in the easy group received a recognition test with 67% old words and 33% new words, whereas the difficult group received a test with 33% old words and 67% new words. Again, the easy group evaluated their performance more favorably than the difficult group, although there were no differences in actual performance. In sum, we presented evidence that processing fluency affects comparative performance evaluations across a wide range of tasks.

A large body of research has shown that people judge themselves to be better than average: They report to be more athletic, better organized, better drivers, better workers, better leaders, fairer, and more polite than others (see Chambers & Windschitl, 2004). Kruger (1999) discussed these findings and found that easy tasks yielded above-average effects and difficult tasks yielded below-average effects. Our studies were not designed to address the issue of above-average versus below-average effects, but a processing fluency approach, as advocated here, can easily accommodate both above-average and below average effects. In a similar vein, we did not address the question whether the indirect method to assess comparative performance judgments would yield the same findings as with the direct method; this issue awaits further research.

Our data suggest that effects of processing fluency on comparative performance evaluations cannot be explained by mood effects. First, differences in well-being could not account for the differences in performance evaluations in the experiments. Effects of mood and well-being on performance-related judgments have been demonstrated (see Reber & Flammer, 2002), and they may play an important role in performance evaluation. However, it cannot explain the effects of fluency on performance evaluation.

The effects of processing fluency on comparative performance evaluations can neither be explained by biases in time or frequency estimates, nor by objective task difficulty. First, neither biased estimates of time nor biased estimates of frequency could account for the differences in performance evaluations in the first three experiments. If participants' memory or estimation of frequencies or times is incorrect, they may well produce biased performance evaluations based on those numbers. However, this mechanism would be different from evaluations based on processing fluency because people then use some fixed numerical value as an anchor for their evaluation. Finally, the performance in the recognition task in Experiment 5 suggested that task difficulty was the same; nevertheless, participants judged the task to be more difficult if only 33% of the items were old words, and performance evaluation was more negative than in the group which got 67% words. In sum, none of the alternative factors can explain the effects of our manipulation of fluency on performance evaluation.

As most judgments, performance evaluations are multidetermined; processing fluency is not the sole determinant of comparative performance evaluations. Further research has to examine when ease of performance is used as information, and when well-being or biased frequency judgments may yield biased performance judgments. Moreover, people may have a fixed standard of comparison, for example, to run the mile within four minutes or to write a chapter within two weeks. If they achieve this goal, they regard it as a success, otherwise as a failure (see Heckhausen, 1991). A social comparison (see Festinger, 1954) is made when people compare themselves to others and conclude that their performance is good or bad, given the distribution of the comparison group. Finally, people may compare their present performance with their performance in the past. This is an individual norm of reference. Such retrospective evaluations of one's own performance are important determinants of expected success or failure in a task, which in turn affects task choice, and persistence in performing the task (see Heckhausen, 1991).

The practical implications of our research are straightforward: People who experience fluent processing during the execution of a task evaluate their performance more positively. The evaluation of one's own performance influences the expected task outcome, which in turn affects the likelihood that people select a task and persist in doing it. Of course, the processes in the sequence from the phenomenal experience of processing fluency to task choice and persistence have to be examined carefully. We have provided evidence that at an early stage, experienced fluency has an impact on the comparative evaluation of performance.

Acknowledgements

This research was supported by the Swiss National Science Foundation (Grant No. 61-57881.99 to Rolf Reber). We thank Christa Thomke, Pascal Wurtz, and Thomas Zimmermann for their help in running the experiments and Robert Bjork, John Dunlosky, and Sara Hutchison for valuable comments.

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130*, 813–838.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117–140.

Gardiner, J. M., Craik, F. I. M., & Bleasdale, F. A. (1973). Retrieval difficulty and subsequent recall. *Memory and Cognition, 1*, 213–216.

Haddock, G., Rothman, A. J., Reber, R., & Schwarz, N. (1999). Forming judgments of attitude certainty, intensity, and importance: The role of subjective experiences. *Personality and Social Psychology Bulletin, 25*, 771–782.

Heckhausen, H. (1991). *Motivation and action*. Heidelberg: Springer.

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 22–34.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.
Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35*, 157–175.
Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639.
Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221–232.
MacLeod, C. M. (1991). Half a century of research on the Stroop effect—An integrative review. *Psychological Bulletin, 109*, 163–203.
Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory and Cognition, 29*, 222–233.
Mednick, S. A., & Mednick, M. T. (1967). *Examiner's manual: Remote associates test*. Boston: Houghton Mifflin.
Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 223–232.
Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease on processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 69–80.
Reber, R. (2004). Availability. In R. Pohl (Ed.), *Cognitive illusions* (pp. 147–163). Hove: Psychology Press.
Reber, R., & Flammer, A. (2002). The development of gender differences in affective expression and in the relationship between mood and achievement-related self-judgments. *European Journal of Psychology of Education*, 377–392.
Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic preference: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8*, 364–382.
Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition, 13*, 47–60.
Rothman, A. J., & Schwarz, N. (1998). Constructing perceptions of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin, 24*, 1053–1064.
Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review, 2*, 87–99.
Schwarz, N. (2002). Feelings as information: Moods influence judgments and processing strategies. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases. The psychology of intuitive judgment* (pp. 534–547). Cambridge: Cambridge University Press.
Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195–202.
Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 907–912.
Simon, D. A., & Bjork, R. A. (2002). Models of performance in learning multisegment movement tasks: Consequences for acquisition, retention, and judgments of learning. *Journal of Experimental Psychology: Applied, 8*, 222–232.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.
Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests: Administration, norms, and commentary* (2nd ed.). New York: Oxford University Press.
Stroop, J. R. (1935). Studies of inference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.
Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235–1253.
Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis of feelings of familiarity. *Acta Psychologica, 98*, 141–166.
Whittlesea, B. W. A., & Williams, L. D. (2000). The source of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 547–565.
Williams, K. W., & Durso, F. T. (1986). Judging category frequency: Automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 387–396.
Winkielman, P., Schwarz, N., & Belli, R. F. (1998). The role of ease of retrieval and attribution in memory judgments: Judging your memory as worse despite recalling more events. *Psychological Science, 9*, 124–126.

Winkielman, P., Schwarz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum.

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological Bulletin, 106*, 231–248.