Davies, Martin F.

Confirmatory bias in the evaluation of personality descriptions: positive test strategies and output interference

Originally published: Journal of Personality and Social Psychology, Vol. 85, No. 4, 2003, 736–744
DOI: 10.1037/0022-3514.85.4.736

Confirmatory Bias in the Evaluation of Personality Descriptions: Positive Test Strategies and Output Interference

Martin F. Davies
University of London

Correspondence concerning this article should be addressed to Martin F. Davies, Department of Psychology, Goldsmiths College, University of London, New Cross, London SE14 6NW, England.
E-mail: psa01mfd@ gold.ac.uk

The operation of confirmatory bias in the endorsement of personality descriptions was examined in 4 studies. Unlike M. F. Davies (1997), who provided only inferential evidence for the role of this bias, the present studies provided direct evidence through the experimental manipulation of supporting versus contradictory cognitions. Generating supporting thoughts resulted in greater acceptance whereas generating contradictory thoughts resulted in lower acceptance of personality descriptions. Supporting cognitions were found to be generated before contradictory cognitions in line with a positive test strategy but evidence was also found for an output interference effect (generating one type of thought interfered with the generation of the opposite type) and it was suggested that confirmatory bias is due to the operation of both mechanisms.

Confirmatory bias is the tendency to seek out information that is consistent with a belief or hypothesis and to ignore or overlook information that is potentially inconsistent (see Evans, 1989; Plous, 1993). Klayman and Ha (1987) proposed that an important aspect of confirmatory bias is the use of a *positive test strategy*. Such a strategy involves testing a hypothesis by looking for examples where the hypothesized property is known or expected to be present rather than absent. For example, in testing the hypothesis that a person is an extravert, a questioner will ask questions such as "Do you like parties?" that, if answered affirmatively, would confirm the hypothesis rather than questions such as "Do you like to be on your own?" that, if answered affirmatively, would disconfirm the hypothesis (e.g., Snyder & Swann, 1978).

In the case of judgments about the self, we possess complex and detailed knowledge of our past history that is diverse and often contradictory (Greenwald & Pratkanis, 1984; Savin-Williams & Demo, 1983). For many aspects of the self, therefore, people should be able to find evidence that confirms a given hypothesis through the use of a positive test strategy. In a series of studies, Davies (1997) found that people use a positive test strategy when evaluating the accuracy of personality feedback. They generated more confirmatory than disconfirmatory cognitions in response to the feedback and their accuracy ratings reflected this relative availability of confirming versus disconfirming evidence. However, this research provided only indirect evidence (through correlational and covariance analyses) for the role of confirmatory bias because confirmatory and disconfirmatory processing were not experimentally manipulated.

In the present series of studies, I set out to experimentally manipulate confirmatory and disconfirmatory processing in the evaluation of self-relevant information. Previous research has shown that people automatically retrieve confirmatory evidence when making judgments. For example, Koriat, Lichtenstein, and Fischhoff (1980) had participants answer general knowledge questions, first choosing which of two alternatives was the correct answer and then estimating the probability that they were correct. Under normal conditions, participants showed typical overconfidence effects, but, when they were asked to generate reasons for and against the possible answers before choosing the correct answer, they showed significantly less overconfidence. Koriat et al. found that it was not the generation of reasons per se that improved judgments but rather the generation of reasons contradicting the chosen answer. When participants generated reasons supporting the chosen answer, Koriat et al. found no improvement in judgments.

In the present series of studies, I investigated people's evaluation of self-relevant information presented in the form of personality statements after they had generated either confirmatory and/or disconfirmatory cognition relating to the statements. Rather than use a personality feedback paradigm (as in Davies, 1997), I used a paradigm in which participants simply evaluated how accurately personality statements described them.

Experiment 1: Confirmatory and Disconfirmatory Processing in the Endorsement of Generalized Personality Descriptions

Method

Participants. Eighty-five college students (31 men and 54 women) were recruited for the study as participants in a project on "understanding personality processes." The mean age of the sample was 21.4 years (*SD* = 3.2; ages ranged from 18 to 32).

Procedure. Participants were given a booklet containing instructions and 13 generalized personality descriptions corresponding to the original statements used by Forer (1949; e.g., "You have a great need for other people to like and admire you."). The statements were described as having been devised for a new personality questionnaire, which the experimenter was interested

in testing on college students. The order of presentation of the statements was randomized across participants.

Following the Cacioppo and Petty (1981) thought-listing procedure, participants in the thought-listing conditions were asked to list their thoughts relating to each statement on a separate page in the booklet, each page being divided into 10 boxes such that each thought was written in a separate box. When evaluating each statement for its self-relevance, they were asked to think of examples from their past history including previous behaviors, thoughts, or feelings. Participants in the *confirmatory* condition (n = 22) were asked to write down examples from their past history in the thought-listing sections of the booklet labeled "Examples *consistent* with the personality statement." Participants in the *disconfirmatory* condition (n = 21) were asked to write down examples from their past history in the thought-listing sections of the booklet labeled "Examples *inconsistent* with the personality statement." Participants in the *confirmatory and disconfirmatory* condition (n = 21) were asked to write down examples from their past history in the thought-listing sections of the booklet labeled "Examples *consistent* and *inconsistent* with the personality statement." They identified whether each example was consistent or inconsistent by writing a *C* or an *I* next to it. Participants worked at their own pace; there was no time limit on completing the thought-listing tasks. (Participants in the *no-thought-listing* condition [n = 21] were not asked to generate or list their thoughts.) After the thought-listing tasks, participants then judged how accurately each statement described them on a scale that ranged from 1 (*not at all accurate*) to 10 (*extremely accurate*). Finally, participants were thoroughly debriefed about the study.

Results

The mean accuracy ratings are shown in Table 1. There was a significant main effect of thought-listing condition, $F(3, 81) = 10.96$, $p < .001$. Planned comparisons showed that accuracy ratings were significantly higher in the confirmatory-only condition (M = 7.70) than in the no-thought condition (M = 7.06), $t(81) = 2.06$, $p < .05$, significantly lower in the disconfirmatory-only condition (M = 5.96) than in the no-thought condition, $t(81) = -3.51$, $p < .001$, with no difference between the both condition (M = 7.15) and the no-thought condition ($t < 1$). Further planned comparisons showed that accuracy ratings were slightly higher in the confirmatory-only condition than in the both condition, $t(81) = 1.75$, $p < .10$, and lower in the disconfirmatory-only condition than in the both condition, $t(81) = -3.81$, $p < .001$.

The numbers of consistent and inconsistent thoughts generated in the different thought-listing conditions are also shown in Table 1. Significantly more consistent thoughts (M = 2.69) than inconsistent thoughts (M = 1.64) were generated in the both condition, $F(1, 20) = 16.57$, $p < .002$, and more consistent thoughts were generated in the confirmatory-only condition (M = 3.76) than inconsistent thoughts in the disconfirmatory-only condition (M = 3.16), $F(1, 41) = 4.58$, $p < .05$. In addition, more consistent thoughts were generated in the confirmatory-only condition than in the both condition, $F(1, 41) = 14.60$, $p < .001$, and more inconsistent thoughts were generated in the disconfirmatory-only condition than in the both condition, $F(1, 40) = 28.62$, $p < .001$.

Discussion

The results of Experiment 1 showed that generating only disconfirmatory thoughts led to significantly lower accuracy ratings for the personality descriptions than generating both types of thought or no thoughts, whereas generating only confirmatory thoughts led to significantly higher accuracy ratings than generating both types of thought or no thoughts. Generating both confirmatory and disconfirmatory thoughts did not produce accuracy ratings different from those in the no-thought condition. These results support the findings of Davies (1997), who found that the more confirmatory thoughts generated the higher the accuracy ratings of personality feedback and the more disconfirmatory thoughts generated the lower the accuracy ratings of personality feedback. The results however only partly support the findings of Koriat et al. (1980)

on the effects of generating supporting versus contradictory reasons on overconfidence in probability judgments. They found that generating supporting reasons did not produce differences in accuracy of probability judgments compared with no generation, whereas, in the present study, generating only confirmatory thoughts did lead to significantly higher accuracy ratings than the no-thought condition. There are a number of differences between the two studies that could account for this discrepancy. The most likely possibility is that in Koriat et al.'s second experiment, participants generated only one reason supporting and/or contradicting the chosen answer, whereas in the present study (as in Davies, 1997) participants could generate as many relevant thoughts as they liked. This may seem like a minor difference, but people's judgments may well have been influenced by the amount of evidence supporting or contradicting a given personality statement. In fact, correlational analyses supported this reasoning. In the confirmatory condition, there was a significant positive correlation between number of thoughts listed and accuracy ratings ($r = .52$, $p < .02$). In the disconfirmatory condition, there was a significant negative correlation between number of thoughts listed and accuracy ratings ($r = -.47$, $p < .05$). In the both condition, there was a significant positive correlation between number of consistent thoughts and accuracy ratings ($r = .43$, $p < .05$) and a significant negative correlation between number of inconsistent thoughts and accuracy ratings ($r = -.46$, $p < .05$). Of course, these correlational findings do not prove that differences in the number of thoughts generated led to differences in perceived accuracy. Not only is it possible that judgments of perceived accuracy could have led to differences in the number of thoughts generated, but there may well be a third unknown factor that accounts for the observed correlations. In addition, it is possible that participants' accuracy ratings were based on ease of generation rather than number of reasons generated in line with Tversky and Kahneman's (1973) analysis of the availability heuristic. The present study unfortunately cannot distinguish between these different possibilities.

Experiment 2: Confirmatory and Disconfirmatory Processing in the Endorsement of True Versus False Personality Descriptions

To increase the generalizability of the findings of Experiment 1, in Experiment 2 I examined the role of confirmatory and disconfirmatory processing in the evaluation of personality descriptions that were derived from actual psychometric test results rather than generalized personality statements. One problem with the use of Barnum statements is that their accuracy is unknown or uncertain. The use of well-validated psychometric instruments provides a better criterion against which to compare participants' assessments of accuracy (although, of course such a criterion is a relative rather than an absolute or objective standard of accuracy).

Method

Participants. One hundred two college students (38 men and 64 women) were recruited for the study as participants in a project on "understanding personality processes." The mean age of the sample was 21.6 years ($SD = 3.4$; ages ranged from 18 to 35).

Procedure. In an initial mass testing session, participants were administered a number of personality inventories and psychometric tests, including the 16 Personality Factor (16PF) Questionnaire (Form A; Cattell, Eber, & Tatsuoka, 1970). For the experimental session approximately 4 weeks later, participants were informed that they would be given personality statements to evaluate. They were not told that the statements were based on their responses to an earlier personality questionnaire (the 16PF), and the results of the postexperimental debriefing did not reveal that any participants guessed that the statements were based on their earlier responses. The statements were described as having been devised for a new personality questionnaire, which the experimenter was interested in testing on college students. The order of presentation of the statements was randomized across participants. Half the participants were given 16 personality statements based on their genuine 16PF scores and the other half were given 16 statements based on false 16PF scores produced by reversing the polarity

of their real scores. The statements consisted of the brief 16PF descriptors with appropriate adverbial qualifiers (sten scores 1,10 = *extremely*; sten scores 2,9 = *very*; sten scores 3,8 = *moderately*; sten scores 4,7 = *somewhat*; sten scores 5,6 = *slightly*).

The thought-listing and accuracy rating tasks were the same as those of Experiment 1. Participants in the confirmatory condition (n = 27) were asked to list thoughts that were consistent with the personality statements. Participants in the disconfirmatory condition (n = 25) were asked to list thoughts that were inconsistent with the personality statements. Participants in the confirmatory and disconfirmatory condition (n = 26) were asked to list thoughts that were consistent and inconsistent with the personality statements. Participants in the no-thought-listing condition (n = 24) were not asked to list their thoughts. After the thought-listing tasks, participants then judged how accurately each statement described them on a scale that ranged from 1 (*not at all accurate*) to 10 (*extremely accurate*). Finally, participants were thoroughly debriefed about the study.

Results

The mean accuracy ratings are shown in Table 2 as a function of thought-listing condition and true/false personality statements.

As expected, there was a significant main effect of true versus false statements, $F(1, 94) = 32.66$, $p < .001$, such that descriptions based on true personality scores were judged significantly more accurate (M = 6.80) than those based on false personality scores (M = 5.40). There was also a significant main effect of thought-listing condition, $F(3, 94) = 6.31$, $p < .001$. Planned comparisons showed that accuracy ratings were significantly higher in the confirmatory-only condition (M = 6.82) than in the no-thought condition (M = 6.10), $t(94) = 2.06$, $p < .05$, significantly lower in the disconfirmatory-only condition (M = 5.33) than in the no-thought condition, $t(94) = -2.20$, $p < .05$, with no difference between the both condition (M = 6.16) and the no-thought condition ($t < 1$). Further planned comparisons showed that accuracy ratings were slightly higher in the confirmatory-only condition than in the both condition, $t(94) = 1.94$, $p < .06$, and lower in the disconfirmatory-only condition than in the both condition, $t(94) = -2.40$, $p < .02$.

The numbers of consistent and inconsistent thoughts generated in the different thought-listing conditions for true and false personality statements are also shown in Table 2. As can be seen, significantly more consistent thoughts (M = 2.28) than inconsistent thoughts (M = 1.96) were generated in the both condition, $F(1, 24) = 5.81$, $p < .05$, and more consistent thoughts were generated in the confirmatory-only condition (M = 2.92) than inconsistent thoughts in the disconfirmatory-only condition (M = 2.63), $F(1, 48) = 4.31$, $p < .05$. In addition, more consistent thoughts were generated in the confirmatory-only condition than in the both condition, $F(1, 49) = 20.98$, $p < .001$, and more inconsistent thoughts were generated in the disconfirmatory-only condition than in the both condition, $F(1, 47) = 27.67$, $p < .001$. There was also a significant interaction of consistent/inconsistent thoughts and true versus false statements. As can be seen from the confirmatory and disconfirmatory conditions, more consistent thoughts (M = 3.38) than inconsistent thoughts (M = 2.16) were generated for true statements, whereas more inconsistent thoughts (M = 3.10) than consistent thoughts (M = 2.45) were generated for false statements, $F(1, 48) = 45.19$, $p < .001$. Also, in the both condition, more consistent (M = 2.52) than inconsistent (M = 1.58) thoughts were generated for true statements whereas more inconsistent (M = 2.34) than consistent (M = 2.03) thoughts were generated for false statements, $F(1, 24) = 22.76$, $p < .001$.

Discussion

These results confirm the findings of Experiment 1. Generating only disconfirmatory thoughts led to significantly lower accuracy ratings for the personality statements than generating both types of thought or no thoughts, whereas generating only confirmatory thoughts led to significantly

higher accuracy ratings than generating both types of thought or no thoughts. Generating both confirmatory and disconfirmatory thoughts did not produce accuracy ratings different from those in the no-thought condition. Although it is true that people tend to spontaneously think of more evidence supporting rather than contradicting a statement or outcome (as shown in the both condition), it is not the case that generating supporting evidence has no effect on their judgments. Unlike Koriat et al. (1980), the results of both experiments presented here show that the amount of supporting or consistent evidence generated does affect judgments. Again, in line with the results of Experiment 1, correlational analyses showed that in the confirmatory-only condition there was a significant positive correlation between number of thoughts listed and accuracy ratings ($r = .48$, $p < .02$), in the disconfirmatory-only condition there was a significant negative correlation between number of thoughts listed and accuracy ratings ($r = -.51$, $p < .01$), and in the both condition there was a significant positive correlation between number of consistent thoughts and accuracy ratings ($r = .57$, $p < .01$) and a significant negative correlation between number of inconsistent thoughts and accuracy ratings ($r = -.48$, $p < .05$). However, as in the *Discussion* of Experiment 1, it must be emphasized that these correlational findings do not prove that differences in the number of thoughts generated led to differences in perceived accuracy. It is possible that judgments of perceived accuracy could have led to differences in the number of thoughts generated or there may have been a third unknown factor that accounts for the observed correlations.

One set of findings that has not been discussed so far is the fact that when required to list only one type of evidence (either consistent or inconsistent thoughts), participants generated more thoughts of a given type than when required to list both types of thought. In Experiment 1, those in the confirmatory-only condition generated 3.76 consistent thoughts whereas those in the both condition generated 2.69 consistent thoughts; those in the disconfirmatory-only condition generated 3.16 inconsistent thoughts whereas those in the both condition generated 1.64 inconsistent thoughts. Similarly, in Experiment 2, those in the confirmatory-only condition generated 2.92 consistent thoughts whereas those in the both condition generated 2.28 consistent thoughts; those in the disconfirmatory-only condition generated 2.63 inconsistent thoughts whereas those in the both condition generated 1.96 inconsistent thoughts.

These findings could simply be due to some self-limiting process whereby only a certain number of total thoughts can be generated in a reasonable time period, so that participants in the both condition generate fewer thoughts of each type. However, inspection of the thought listings in the both condition revealed that participants tended to list supporting or consistent evidence before listing contradictory or inconsistent evidence. It is well known that in the process of making judgments or answering questions, people naturally tend to recruit information consistent with their tentatively preferred decision or answer (e.g., Dellarosa & Bourne, 1984; Koriat et al., 1980). This leads to a positive test strategy and confirmatory bias. What is less well known is that evidence generation is subject to interference effects (e.g., Hoch, 1984). In research on memory, it has been found that the very act of recall produces interference and forgetting (Roediger, 1978). For example, Rundus (1973) found that previously recalled list items interfered with the recall of new, unrecalled list items. According to Rundus, output interference is due in part to the strengthening of associations between the retrieval cues and the recalled items produced by the act of recall. This strengthening of associations increases the likelihood of retrieving the previously recalled items but at the expense of the new unrecalled items. By analogy with these memory findings, Hoch (1984) argued that generating a given reason for a decision or judgment will reduce the availability of other reasons through output interference. In particular, generating a supporting reason should interfere with the ability to generate a contradictory reason, and vice versa. In three experiments on predictive judgment, Hoch provided confirmation of output interference by manipulating the order in which supporting (pro) and contradictory (con) reasons were generated. He found that pro–con generation produced not only more pro than con reasons but also higher likelihood judgments than con–pro generation.

Experiment 3: Order of Generating Confirmatory and Disconfirmatory Cognitions in Evaluating Personality Descriptions

In the evaluation of self-relevant information, people may be subject to similar interference effects: The generation of supporting evidence inhibits the generation of contradictory evidence and vice versa. This may be why, in Experiments 1 and 2, participants generated more supporting cognitions in the confirmatory-only condition and more contradictory cognitions in the disconfirmatory-only condition than those in the both condition did. In Experiment 3, I tested this output interference explanation by having some participants first generate a set of thoughts consistent with a personality description and then a set of thoughts inconsistent with the description (pro–con order) while other participants generated thoughts in the reverse order (con–pro). I expected that more thoughts would be generated for the first set than for the second set and that this order of generation effect would be associated with higher acceptance of the personality descriptions in the pro–con order than in the con–pro order. In comparison conditions, participants either generated thoughts in any order or generated no thoughts.

Method

Participants. Ninety-one college students (26 men and 65 women) were recruited for the study as participants in a project on "understanding personality processes." The mean age of the sample was 21.1 years (*SD* = 3.1; ages ranged from 18 to 31).

Procedure. Participants were given a booklet containing instructions and the 13 generalized personality descriptions corresponding to the original Barnum statements used by Forer (1949). The statements were described as having been devised for a new personality questionnaire, which the experimenter was interested in testing on college students. The order of presentation of the statements was randomized across participants. The thought-listing and accuracy rating tasks were the same as those used in Experiment 1. Participants in the *pro–con order* condition (*n* = 22) were asked first to list thoughts that were consistent with the personality statements and then thoughts that were inconsistent with the personality statements. Participants in the *con–pro order* condition (*n* = 21) were asked first to list thoughts that were inconsistent with the personality statements and then thoughts that were consistent with the personality statements. Participants in the *any-order* condition (*n* = 24) were asked to list thoughts that were consistent and inconsistent with the personality statements. Participants in the no-thought-listing condition (*n* = 24) were not asked to list their thoughts. After the thought-listing task, participants then judged how accurately each statement described them on a scale that ranged from 1 (*not at all accurate*) to 10 (*extremely accurate*). Finally, participants were thoroughly debriefed about the study.

Results

The mean accuracy ratings are shown in Table 3. There was a significant main effect of thought-listing condition, $F(3, 87) = 4.91$, $p < .005$. Planned comparisons showed that accuracy ratings were not significantly higher in the pro–con order condition ($M = 7.29$) than in the no-thought condition ($M = 7.10$; $t < 1$), but were significantly lower in the con–pro order condition ($M = 6.19$) than in the no-thought condition, $t(87) = -2.85$, $p < .005$, with no difference between the any-order condition ($M = 7.21$) and the no-thought condition ($t < 1$). Further planned comparisons showed that accuracy ratings were not significantly higher in the pro–con order condition than in the any-order condition ($t = 1$), but were significantly lower in the con–pro order condition than in the any-order condition, $t(87) = -3.18$, $p < .005$. The numbers of consistent and inconsistent thoughts generated in the different thought-listing conditions are also shown in Table 3. Significantly more consistent thoughts ($M = 2.63$) than inconsistent thoughts ($M = 2.13$) were generated, $F(1, 64) = 9.28$, $p < .005$. However, this effect was qualified by a significant interaction

with thought-listing condition, $F(1, 64) = 29.30$, $p < .001$. As can be seen, significantly fewer consistent thoughts were generated in the con–pro order than in the pro–con or any-order conditions, $t(64) = 4.65$, $p < .001$, whereas there was no difference between the pro–con and any-order condition, $t(64) = 1.39$, *ns*. Significantly more inconsistent thoughts were generated in the con–pro order than in the pro–con or any-order conditions, $t(64) = 5.53$, $p < .001$, whereas there was no difference between the pro–con and any-order condition ($t < 1$).

There was a significant positive correlation between accuracy ratings and numbers of consistent thoughts ($r = .44$, $p < .001$) and a significant negative correlation between accuracy ratings and numbers of inconsistent thoughts ($r = -.54$, $p < .001$). To investigate the mediating role of thought generation, an analysis of covariance (ANCOVA) was carried out. Such an analysis can provide evidence for the causal role of thought generation in judgments (e.g., Hoch, 1984). Using consistent and inconsistent thoughts as covariates in an ANCOVA of the accuracy ratings, it was found that the thought covariates were highly significant, $F(2, 62) = 11.12$, $p < .001$. Consequently, the significant effect of thought-listing condition on accuracy ratings obtained in the equivalent analysis of variance (ANOVA), $F(2, 64) = 7.10$, $p < .01$, was reduced to nonsignificance ($F < 1$) by the presence of the covariates.

Discussion

The present findings demonstrated output interference effects in the generation of thoughts and in rating the accuracy of the personality descriptions. In the pro–con order condition, significantly more consistent and fewer inconsistent thoughts were generated and accuracy ratings were significantly higher than in the con–pro order condition. However, the pro–con order condition did not differ significantly from the any-order condition or the no-thoughts condition. This confirms the impression from the previous experiments that, when generating both consistent and inconsistent thoughts, participants tended to produce supporting cognitions before contradictory cognitions. Formal analyses of the order of thoughts generated in the both/any-order conditions of the three studies showed that the first thoughts generated were significantly more likely to be consistent than inconsistent. The percentage of first thoughts that were consistent were as follows: Experiment 1 = 64% (Binomial test $p = .007$), Experiment 2 = 65% (Binomial test $p = .003$), and Experiment 3 = 74% (Binomial test $p < .001$).

These results also mean that, when spontaneously generating thoughts, there is an automatic output interference effect leading to more consistent than inconsistent thoughts and higher accuracy ratings for personality descriptions. Not only does the use of a positive test strategy lead to participants seeking out confirmatory evidence, but this in turn inhibits the retrieval of disconfirmatory evidence. By contrast, when participants are required to generate contradictory cognitions first, there is a dramatic effect on the generation of consistent and inconsistent thoughts and on ratings of accuracy.

There is however an alternative explanation for the findings of Experiment 3. Rather than showing an output interference effect, the findings may reflect some sort of fatigue effect. That is, after generating the first type of thoughts, participants get tired and generate fewer of the second type of thoughts. Hoch (1984) considered the possibility that the output interference he had observed in the generation of reasons was due to fatigue or boredom. He argued that this rival explanation should lead to a decreasing trend in the number of reasons generated for each successive generation task. However, he found no evidence for such a decreasing trend across trials. I too carried out similar analyses on the number of consistent and inconsistent thoughts generated across trials (the 13 statements) and found no evidence for significant trends contrary to a fatigue explanation.

However, even though no decreasing trend in number of thoughts generated was found across trials, it could still be that within each trial (i.e., for each statement) participants lose motivation or ability to generate more and more thoughts (e.g., Dijksterhuis, Macrae, & Haddock, 1999; Schwarz et al., 1991). To test this possibility, in Experiment 4, I used an experimenter

paced procedure for thought listing so that the generation of the first type of thoughts and the second type of thoughts occurred in separate time periods unlike in the previous studies.

Experiment 4: Output Interference Effects and Personality Questionnaire Responding

In this experiment, I sought to rule out a motivation/ability decrement explanation for the output interference effects observed in Experiment 3 by having the first and second types of thought generated in separate time periods. I also sought to extend the findings further to the prototypical paradigm of personality questionnaire responding (in this particular case, filling in the Neuroticism scale of the Eysenck Personality Questionnaire [EPQ]). I expected more consistent than inconsistent thoughts to be generated in the pro–con order of generation than in the con–pro order and that Neuroticism scores would consequently be higher in the pro–con than in the con–pro order.

Method

Participants. Eighty-eight college students (24 men and 64 women) were recruited for the study as participants in a project on "understanding personality processes." The mean age of the sample was 21.3 years (*SD* = 3.3; ages ranged from 18 to 37).

Procedure. Participants were given a booklet containing instructions and the 12 questions of the Neuroticism scale of the short form Eysenck Personality Questionnaire–Revised (Eysenck & Eysenck, 1975). The questions were described as having been devised for a new personality questionnaire, which the experimenter was interested in testing on college students. The order of presentation of the questions was randomized across participants. Participants in the pro–con order condition (n = 30) were asked first to list thoughts that were consistent with a "yes" answer to the questions and then thoughts that were consistent with a "no" answer to the questions. Participants in the con–pro order condition (n = 29) were asked first to list thoughts that were consistent with a "no" answer to the questions and then thoughts that were consistent with a "yes" answer to the questions. Participants were told that the study was experimenter paced and that they should turn to the next page only when instructed. They had 90 s to write down their first set of thoughts and then 90 s to write down their second set of thoughts. Participants in the no-thought-listing condition (n = 29) were not asked to list their thoughts. After the thought-listing tasks, participants then gave their judgments as to whether each question described them or not (yes/no). All the items of the EPQ Neuroticism scale are phrased in a positive direction (yes = 1, no = 0) so that higher scores represent higher neuroticism. Finally, participants were thoroughly debriefed about the study.

Results and Discussion

The mean Neuroticism scores are shown in Table 4. There was a significant main effect of thought-listing condition, $F(2, 85) = 4.39$, $p < .02$. Planned comparisons showed that Neuroticism scores were not significantly higher in the pro–con order condition (M = 6.33) than in the no-thought condition (M = 5.97; $t < 1$), but were significantly lower in the con–pro order condition (M = 4.69) than in the no-thought condition, $t(85) = -2.18$, $p < .05$.

The numbers of "yes" consistent and "no" consistent thoughts generated in the different thought-listing conditions are also shown in Table 4. Overall, significantly more "yes" consistent thoughts (M = 2.65) than "no" consistent thoughts (M = 2.21) were generated, $F(1, 57) = 4.51$, $p < .05$. However, this effect was qualified by a significant interaction with thought-listing condition, $F(1, 57) = 19.36$, $p < .001$. As can be seen, significantly more "yes" consistent thoughts were generated in the pro–con order condition (M = 3.19) than in the con–pro order condition (M = 2.10), simple effect $F(1, 57) = 20.81$, $p < .001$, whereas significantly more "no" consistent

thoughts were generated in the con–pro order condition (M = 2.56) than in the pro–con order condition (M = 1.87), simple effect $F(1, 57) = 5.75$, $p < .02$.

There was a significant positive correlation between Neuroticism scores and numbers of "yes" consistent thoughts ($r = .46$, $p < .001$) and a significant negative correlation between Neuroticism scores and numbers of "no" consistent thoughts ($r = -.51$, $p < .001$). Using "yes" and "no" consistent thoughts as covariates in an ANCOVA of the Neuroticism scores, it was found that the thought covariates were highly significant, $F(2, 55) = 10.78$, $p < .001$. Consequently, the significant effect of thought-listing condition on Neuroticism scores obtained in the equivalent ANOVA, $F(1, 57) = 7.97$, $p < .01$, was reduced to nonsignificance ($F < 1$) by the presence of the covariates. These results therefore confirm the findings of Experiment 3 and support the existence of an output interference effect when people respond to personality questionnaire items.

General Discussion

Confirmatory bias is the tendency to seek out information that is consistent with a belief or hypothesis and to ignore or overlook information that is inconsistent. A key element in confirmatory bias is the use of a positive test strategy. Such a strategy involves testing a hypothesis by looking for examples where the hypothesized property is expected to be present rather than absent. This positive test strategy is likely to lead to hypothesis confirmation if evidence consistent with the hypothesis is obtained. For example, Kunda, Fong, Sanitioso, and Reber (1993) found that when asked directional questions such as "Are you happy with your social life?" or "Are you extraverted?" respondents tended to think of examples confirming rather than contradicting the hypothesized characteristic and they concluded that they were more likely to possess the attribute than those respondents asked the opposite question.

Davies (1997) found that people use a positive test strategy when evaluating the accuracy of personality feedback. They generated more confirmatory than disconfirmatory cognitions in response to the feedback and their accuracy ratings reflected this relative availability of confirming versus disconfirming evidence. However, this research and the research by Kunda et al. (1993) provided only inferential evidence for the role of positive test strategies because confirmatory and disconfirmatory processing were not experimentally manipulated.

In the present series of experiments, more direct evidence for the role of positive test strategies was obtained by experimentally manipulating confirmatory and disconfirmatory processing. Inducing respondents to generate only thoughts consistent with personality descriptions resulted in more confirmatory cognitions and higher accuracy ratings for the descriptions, whereas inducing respondents to generate only thoughts inconsistent with the personality descriptions resulted in more disconfirmatory cognitions and lower accuracy ratings for the descriptions. Of importance, these findings were obtained outside of the personality feedback paradigm. In the present studies, participants evaluated the extent to which self-relevant statements described them without any suggestion that the statements represented personality feedback. Indeed, in Experiment 4, the participants were responding to genuine personality questionnaire items.

The novel finding arising from Experiments 3 and 4 is the operation of an output interference effect in the evaluation of self-relevant information. Use of a positive test strategy leads people to seek out evidence that is consistent with a hypothesized property. On top of this, however, output interference means that this retrieval of confirming evidence actively inhibits the retrieval of disconfirming evidence further magnifying any confirmatory bias. Thus, when evaluating a self-relevant characteristic, people spontaneously retrieve examples confirming the presence of the hypothesized property (positive test strategy), which in turn makes it more difficult for them to recruit contradictory examples (output interference).

There are a number of significant implications of these findings. The first is the issue of the stability versus malleability of the self concept. At any given time, the "working self concept" (Markus & Kunda, 1986) consists of only a subset of the total store of knowledge about the self.

Any factors or situations that affect the accessibility of confirming or disconfirming information about the self may at least temporarily influence one's view of the self by altering the contents of the working self concept (e.g., Fazio, Effrein, & Falender, 1981; Kunda et al., 1993; Rhodewalt & Agustsdottir, 1986). For example, a person applying for a job as a firefighter would be advised to present themselves as fit, resourceful, brave, yet careful. In communicating this impression to the job interviewer, this person's working self concept would contain many examples that confirm and few that disconfirm these attributes and their postinterview self concept might well differ from their preinterview self concept.

The second implication of the findings is the issue of question phrasing in questionnaires (e.g., Schwarz & Sudman, 1992). If people tend to retrieve confirming evidence for a statement or question, then questionnaires need to be "balanced," such that half the items are phrased in one direction and the other half are phrased in the reverse direction so that simply agreeing with the statements does not automatically lead to a high score on the measure of interest. This is the old issue of acquiescence that is well known but not that well understood (e.g., Paulhus, 1991). However, the present findings go beyond a simple acquiescent response set. They suggest that participants who are motivated to confirm a particular characteristic in the self concept are likely to do so regardless of whether the items are phrased in one direction or the other. What personality or situational factors influence the tendency to confirm or deny a particular trait in the self is an interesting question that deserves further research.

References

Cacioppo, J. T., & Petty, R. E. (1981). Social psychological procedures for cognitive response assessment: The thought-listing technique. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 309–342). New York: Guilford Press.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). *Handbook for the Sixteen Personality Factor Questionnaire*. Champaign, IL: Institute for Personality and Ability Testing.

Davies, M. F. (1997). Positive test strategies and confirmatory retrieval processes in the evaluation of personality feedback. *Journal of Personality and Social Psychology, 73,* 574–583.

Dellarosa, D., & Bourne, L. E. (1984). Decision and memory: Differential retrievability of consistent and contradictory evidence. *Journal of Verbal Learning and Verbal Behavior, 23,* 669–682.

Dijksterhuis, A., Macrae, C. N., & Haddock, G. (1999). When recollective experiences matter: Subjective ease of retrieval and stereotyping. *Personality and Social Psychology Bulletin, 25,* 760–768.

Evans, J. St. B. T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ: Erlbaum.

Eysenck, H. J., & Eysenck, S. (1975). *Manual of the Eysenck Personality Questionnaire*. London: Hodder & Stoughton.

Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perception following social interaction. *Journal of Personality and Social Psychology, 42,* 232–242.

Forer, B. R. (1949). The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology, 44,* 118–123.

Greenwald, A. G., & Pratkanis, A. R. (1984). The self. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 129–178). Hillsdale, NJ: Erlbaum.

Hoch, S. J. (1984). Availability and interference in predictive judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 649–662.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 107–118.

Kunda, Z., Fong, G. T., Sanitioso, R., & Reber, E. (1993). Directional questions direct self-conceptions. *Journal of Experimental Social Psychology, 29,* 63–86.

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self concept. *Journal of Personality and Social Psychology, 51,* 858–866.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 5–32). New York: Academic Press.

Plous, S. (1993). *The psychology of judgment and decision making*. New York: McGraw-Hill.

Rhodewalt, F., & Agustsdottir, S. (1986). Effects of self-presentation on the phenomenal self. *Journal of Personality and Social Psychology, 50,* 47–55.

Roediger, H. L. (1978). Recall as a self-limiting process. *Memory and Cognition, 6,* 54–63.

Rundus, D. (1973). Negative effects of using list items as recall cues. *Journal of Verbal Learning and Verbal Behavior, 12,* 43–50.

Savin-Williams, R. C., & Demo, P. (1983). Situational and transituational determinants of adolescent self-feelings. *Journal of Personality and Social Psychology, 44,* 820–833.

Schwarz, N., & Sudman, S. (Eds.). (1992). *Context effects in social and psychological research*. New York: Springer Verlag.

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.

Snyder, M., & Swann, W. B. (1978). Hypothesis testing processes in social interaction. *Journal of Personality and Social Psychology, 36,* 1202– 1212.

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

Received August 2, 2001
Revision received February 28, 2003
Accepted February 28, 2003 _

Table 1
Accuracy Ratings and Number of Thoughts Generated as a Function of Thought-Listing Condition

Dependent measure	Thought-listing condition			
	No thoughts (n = 21)	Confirmatory (n = 22)	Disconfirmatory (n = 21)	Both (n = 21)
Rated accuracy				
M	7.06	7.70	5.96	7.15
SD	1.09	0.93	1.01	1.05
Consistent thoughts				
M	—	3.76	—	2.69
SD	—	0.95	—	0.89
Inconsistent thoughts				
M	—	—	3.16	1.64
SD	—	—	0.88	0.96

Note. Dashes indicate that no scores or measures were taken.

Table 2
Accuracy Ratings and Number of Thoughts Generated as a Function of Thought-Listing Condition and True/False Personality Statements

Dependent measure	Thought-listing condition			
	No thoughts (n = 24)	Confirmatory (n = 27)	Disconfirmatory (n = 25)	Both (n = 26)
	True statements			
Rated accuracy				
M	6.68	7.59	6.06	6.85
SD	1.33	1.37	1.22	1.23
Consistent thoughts				
M	—	3.38	—	2.52
SD	—	0.59	—	0.53
Inconsistent thoughts				
M	—	—	2.16	1.58
SD	—	—	0.42	0.48
	False statements			
Rated accuracy				
M	5.52	6.04	4.59	5.46
SD	1.20	1.29	1.06	1.10
Consistent thoughts				
M	—	2.45	—	2.03
SD	—	0.47	—	0.43
Inconsistent thoughts				
M	—	—	3.10	2.34
SD	—	—	0.49	0.42

Note. Dashes indicate that no scores or measures were taken.

Table 3
Accuracy Ratings and Number of Thoughts Generated as a Function of Thought-Listing Condition

	Thought-listing condition			
Dependent measure	No thoughts (n = 24)	Pro–con order (n = 22)	Con–pro order (n = 21)	Any order (n = 24)
Rated accuracy				
M	7.10	7.29	6.19	7.21
SD	1.06	1.05	1.16	0.98
Consistent thoughts				
M	—	3.17	1.88	2.80
SD	—	0.87	0.95	0.89
Inconsistent thoughts				
M	—	1.61	3.07	1.78
SD	—	0.90	0.82	1.07

Note. Dashes indicate that no scores or measures were taken.

Table 4
Neuroticism Scores and Number of Thoughts Generated as a Function of Thought-Listing Condition

	Thought-listing condition		
Dependent measure	No thoughts (n = 29)	Pro–con order (n = 30)	Con–pro order (n = 29)
Neuroticism score			
M	5.97	6.33	4.69
SD	2.21	2.11	2.36
"Yes" consistent thoughts			
M	—	3.19	2.10
SD	—	0.86	0.98
"No" consistent thoughts			
M	—	1.87	2.56
SD	—	0.99	1.22

Note. Dashes indicate that no scores or measures were taken.