Overconfidence in Predictions as an Effect of Desirability Bias

F. Giardini(✉), G. Coricelli, M. Joffily, and A. Sirigu

Keywords: Decision making · Overconfidence · Reward · Desirability · Accuracy

1 Introduction

Most people hold unrealistic positive beliefs about their personal skills, their knowledge (Fischoff, Slovic, & Lichtenstein, 1977), and their possibilities to overcome the performance of other individuals (Weinstein, 1980). This general tendency, called *overconfidence*, is a stable and pervasive finding both in many real-life domains and in several experimental settings. People are overconfident about their driving skills (Svenson, 1981), about their ability as basketball players (McGraw, Mellers, & Ritov, 2004), about their competence in financial and managerial problems (Camerer & Lovallo, 1999; Mahajan, 1992), and about their general knowledge (Juslin, 1994; Harvey, 1997). This systematic overestimation of one's own capabilities and probabilities of success can have important consequences, and sometimes results in suboptimal decisions.

While the existence of overconfidence is uncontroversial, its sources and determinants are still open to debate (Ayton & McClelland, 1997; Klayaman, Soll, Gonzalez-Vallejo, & Barlas, 1999).

In this study we contribute to this debate by demonstrating that overconfidence in predictions is related to the desirability of the predicted outcome. When people are required to forecast possible future events, they tend to be more confident in the occurrence of favourable events, with little or no regard for their objective likelihood.

F. Giardini
Institute of Cognitive Sciences and Technologies (ISTC), National Research Council, Via San Martino della Battaglia, 44, 00185, Rome, Italy
e-mail: francesca.giardini@istc.cnr.it

M. Abdellaoui, J.D. Hey (eds.) *Advances in Decision Making Under Risk and Uncertainty*. Theory and Decision Library C 42.

We claim that forecasts are related to the desirability of the evaluated/predicted event, i.e. the more desirable the event, the stronger the belief that this event will happen. By manipulating the level of reward for the correct answer in a visual perceptual task we were able to highlight the presence and the principal characteristics of a *desirability bias*, which affects people's confidence. In our experiments we found a general increase in subjects' confidence levels under a reward versus a no-reward condition. Furthermore, the outcome desirability in terms of relative reward value biased subjects' confidence, leading them to believe that they were more accurate than they actually were.

In what follows we present two studies showing how a desirable result, i.e. a monetary reward, can bias people's confidence judgements in their perceptual accuracy, inducing them to be overconfident. Calibration studies have long investigated people's ability to match their judgements of the relative frequency of an event to the actual likelihood of that event. Perfect calibration occurs when average confidence is equal to the actual frequency of that event, and people are said to be "well calibrated". Unfortunately, this happens quite rarely. Several studies have shown that people are usually poorly calibrated, exhibiting either under- or overconfidence.

Overconfidence is the positive difference between mean reported confidence in the chosen answer and the percentage of correct answers (CONF – % correct answers >0). This phenomenon is preponderant in general knowledge or cognitive tasks (Brenner, Koheler, Liberman, & Tversky, 1996; Fischoff et al., 1977; Klayman et al., 1999; Koriat, Lichtenstein, & Fischhoff, 1980). The inverse phenomenon is underconfidence (CONF – % correct answers <0), which is more frequent in perceptual or in very easy cognitive tasks (Bjorkman, Juslin, & Winman, 1993; Juslin & Olsson, 1997; Keren, 1988).

Therefore, people usually overestimate their ability or knowledge in cognitive tasks, and underestimate the accuracy of their perceptions in sensory tasks.

1.1 Overview

This paper explores the effects of a desirable outcome on people's accuracy and confidence in a visual perceptual task. Effects of motivation on perception, judgement and decision making are well documented, but these effects usually refer to probability evaluation. For instance, the possibility of gaining money induces people to neglect or underestimate the base rate probabilities of events (Bar-Hillel, 1980; Kahneman & Tversky, 1996). Thus the perceived probability of a given event increases as a function of reward, even though the probabilities of success (base rates) are unchanged. A study by Ginossar & Trope (1987) suggested that goals may affect the use of base rate information, and there is some general evidence that motivation may affect the use of statistical heuristics. Generally speaking, the effects of goals and desires on reasoning, forecasting and memory are well documented (for a review, see Kunda, 1990), but less is known about how desirability affects people's confidence.

Here, we will investigate the effect of reward on people's confidence, that is, the degree of belief in a given hypothesis, judgement, or prediction. We claim that a desirable result makes people feel more confident in the possibility of getting it, compared with a neutral outcome, which is neither beneficial nor harmful to them. We call this phenomenon the *desirability bias* and we predict that it will induce individuals to be more confident when the possible reward is higher, all other things being equal. The desirability bias is a motivational effect working on the belief people hold about the likelihood of a certain outcome, and it should be independent from other effects, such as the difficulty of the task.

In Study 1 we tested the effect of three reward levels on confidence judgements in a perceptual task with fixed difficulty. In Study 2 we investigated how confidence judgements vary as a function of reward (low or high) for three levels of difficulty of the task. Manipulating the complexity of the task we induced three levels of accuracy: *Difficult* (Accuracy ≤ 0.5), *Intermediate* ($0.5 <$ Accuracy < 0.75) and *Easy* (Accuracy > 0.75). Along with the reward groups we also tested one (Study 1) and three (Study 2) Control Groups, which performed the same task, with the same difficulty levels, but with no monetary incentives during the experiment. Control groups allow us to set base rate confidence and accuracy levels that are then compared with reward conditions.

We used a perceptual task in order to isolate the effect of motivation, and to exclude other possible explanations for overconfidence, such as failure to think of reasons why one might be wrong (Koriat et al., 1980) or individual's failure to assess the credibility or weight of the evidence (Griffin & Tversky, 1992). The task we implemented has three main characteristics. First, it is divided in two independent parts: the first part is constituted by a low-level perceptual task, requiring no reasoning and cognitive processing, while the second part requires an inferential process to evaluate the reward and to assess the confidence in the performance. Second, it makes it possible to directly correlate subjects' performance, i.e. accuracy of their responses, to reward and to variations in confidence, excluding any sort of other motivational effect, since the reward is displayed only when the perceptual part is over. Finally, the absence of feedback and the controlled number of trials allow us to rule out any kind of learning during the experiment.

2 Study 1

2.1 Participants

Twenty-seven undergraduate students (15 female and 12 male) were recruited to take part in a study at the Experimental Economics Laboratory (LabSi) of the University of Siena. All subjects were naïve with respect to the nature and aims of the experiment. Mean age of participants was 22 years (s.d. = 1.83).

2.2 Stimuli

We used a visual motion discrimination task typically used in neuro-physiological studies with monkeys (Britten, Shadlen, Newsome, & Movshon, 1992; Celebrini & Newsome, 1994). The stimulus display consisted of one white circle on a black background, containing 2000 black dots jiggling toward the right and the left side of the screen, with a fixed percentage of them coherently shifting toward either the left or the right (see Fig. 1).

The difficulty level was determined as the ratio between the velocity of the jiggling movement of the dots in the background and that of the linear movement toward one direction of the set that participants had to identify. We assessed this level in a previous pilot study, where we singled out a level of accuracy (i.e. percentage of correct answers) around 70%.

The background movement consisted of some dots jiggling in a random manner toward the right and some others toward the left, whereas the coherent set consisted of a fixed percentage of dots moving coherently toward only one direction.

Subjects had to identify the "coherent direction" of the dots, separating the coherently moving dots from the background movement. Each of the two directions was equally probable. The stimulus difficulty was set at the beginning of the experiment and then it was kept constant throughout the experimental session. Each stimulus was presented for 2 s.

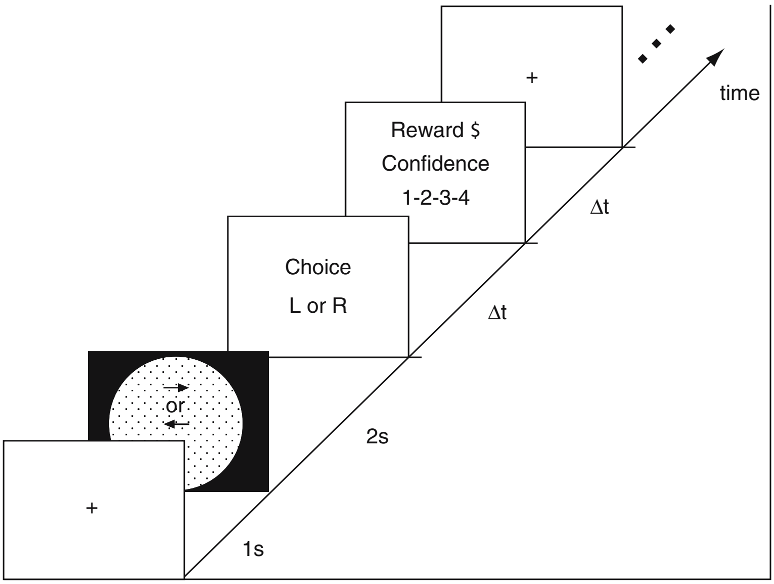

Fig. 1 Time course of the experiment in Study 1: fixation point (1 s), stimulus presentation (2 s), direction selection (left or right; self-paced), reward information (1, 5, or 15 €), and confidence scale (self-paced)

2.3 Procedure

Two groups of subjects were tested in two separate sessions. One group of eighteen subjects participated in a *reward condition* and the second group of nine subjects (control) took part in a *no-reward condition*. Participants had to discriminate the direction of moving dots showed on a computer screen placed in front of each of them.

Each trial began with a fixation point lasting 1 s. which directed the subject's attention toward the centre of the screen, where the stimulus was going to appear. The fixation point was followed by the stimulus presentation (2 s). At the end of the stimulus presentation, the participants saw on the screen the following question: "What was the dots' direction?", and below "Left or Right". They chose (self-paced choice) by pressing the corresponding arrow on the PC keyboard. Once one of the two arrows was pressed they could not modify their choice.

This was followed by a blank screen and then both the reward amount and a confidence scale appeared. All three amounts of reward (1, 5 and 15 €) were equally probable, and subjects were instructed that the computer program randomly paired rewards with stimuli. The uncertainty level for the stimulus recognition was always the same and no correlation existed among reward amounts and stimuli. This was explicitly stated in the Instructions and reminded to the subjects at the beginning of the experimental session.

The reward was showed in the upper part of the monitor (If you detected the correct direction you could be rewarded with … Euro), while in the lower part appeared the question about the degree of confidence (How confident do you feel you detected the correct direction?). Confidence was measured on a 4-points confidence scale ranging from 1. "Not sure at all" to 4. "Really sure", with two intermediate values (2. "Not so sure" and 3. "Sure enough"). Subjects used left and right arrows on the keyboard to state their confidence and they could modify their choice until they pressed "Enter" to confirm it. There was no time limit for reporting confidence level.

Once they reported their confidence they pressed "Enter" to go to the next trial (Fig. 1). No feedback was provided to subjects, neither about the correct direction nor about their winnings. To summarize, the time course of the task was: [fixation point → stimulus → direction choice → possible reward → confidence judgement]→ [fixation point → (…)].

Subjects performed 57 trials, 9 of which were training trials aimed to get them familiarized with the task, while the remaining 48 were experimental trials. There was no time limit for completing the task. At the end of the session the subjects in the reward condition were paid accordingly to their performance in one trial randomly drawn by the computer out of the 48 trials (the 9 training trials were excluded). If their response in the drawn trial was correct they were paid accordingly to the reward shown during that trial, otherwise they only received the participation fee (3 €). These features of the experiment were properly explained in the Instructions subjects read before starting the session.

In a separate session, a Control Group, recruited following the usual procedure of the LabSi, was assigned (randomly) to the no-reward condition and performed an identical task (9 training trials and 48 experimental trials) except for the reward, which was neither mentioned nor displayed. Subjects in this condition were simply asked to individuate the direction and to assess their confidence on the 4-points scale, and received a participation fee of 3 €. Each experimental session lasted approximately 30 min.

2.4 Results

We found that confidence judgements of the correct answer vary with the amount of monetary reward (data from reward condition; repeated measures ANOVA, $F_{(2,17)} = 6.74$, $P = 0.0034$). Confidence increased as reward increased (as shown in Fig. 2), with significant differences between 1 and 15 € (Wilcoxon signed-rank test $z = -2.268$, $P = 0.0233$), and between 5 and 15 € (signed-rank test $z = -2.686$, $P = 0.0072$). No statistically significant difference was found between 1 and 5 € (signed-rank test $z = -0.982$, $P = 0.362$).

Regression analysis (Order Probit model, Table 1) can help us to understand where and how the reward enters the process.

We considered confidence (as reported in the confidence scale, i.e. takes values 1–4) as a function of reaction time (RT, equal to the response time of subject's choice of the direction of the moving dots), accuracy (A, equal to 1 if correct and 0 if incorrect), and reward level ($, only in reward condition, i.e. takes values of 1, 5 or 15). In the control group, which did not receive any reward, confidence level was a function of accuracy (A) and Reaction Time (RT). The time taken to respond as well as the accuracy of the responses determined subjects' confidence judgements when they did not have the possibility of getting any monetary reward

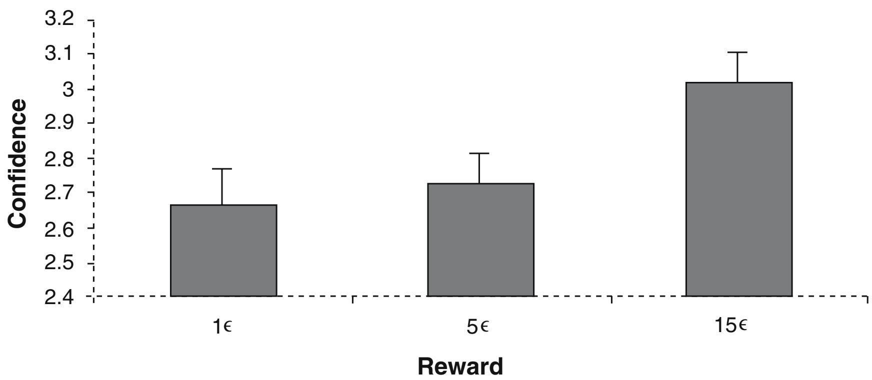

Fig. 2 Mean confidence (+standard errors) for the three different reward levels. Confidence was significantly higher for 15 compared with 1 and 5 €

Table 1 Study 1 regression analysis. Confidence levels as a function of Accuracy, reward level (only in reward condition) and reaction time. Regression analysis (Order Probit). The dependent variable Confidence takes values of 1–4; A = accuracy, equal to 1 if correct and 0 if incorrect; $ = reward level (i.e. 1, 5, or 15); RT = reaction time

Regression analysis order probit: dependent variable is "*CONFIDENCE*"				
a. *Data from experimental sessions ($)*				
Variable	Coeff.	Std. error	Z	P > |z|
A	0.133	0.082	1.64	0.102
$	0.035	0.006	5.51	0.000
RT	−0.0003	0.000	−7.67	0.000
Number of obs = 864				
Log likelihood = -997.39701				
Prob > Chi^2 = 0.0000				
b. *Data from control sessions (no $)*				
Variable	Coeff.	Std. error	Z	P > |z|
A	0.336	0.118	2.84	0.004
RT	−0.0002	0.000	−3.77	0.000
Number of obs = 432				
Log likelihood = −500.13509 Prob > Chi^2 = 0.0000				

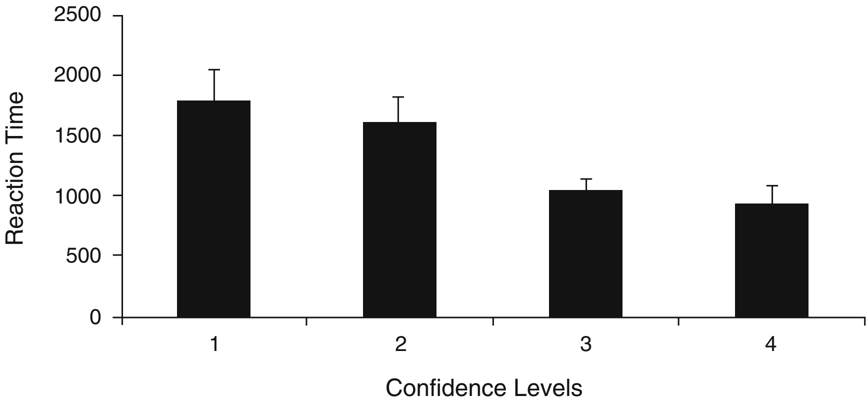

Fig. 3 Mean reaction time (RT; data collapsed over reward and no-reward conditions) for each level of confidence judgement of correct answer (with 1 – "Not sure at all" and 4 – "Really sure"). RT was inversely correlated with Confidence level

for a correct answer. By contrast, in the experimental condition, reward ($) and RT were significantly correlated with the confidence level, whereas accuracy was not. The inverse relationship between RT and confidence present in both conditions is shown in Fig. 3. The mean accuracy was not significantly different between the reward (Mean = 0.71) and the no-reward (Mean = 0.73) conditions (two-sample Wilcoxon rank-sum test $z = -0.258$, $P = 0.79$).

Thus, the presence of a monetary reward biases individuals' confidence, no matter how accurate they have been. That is, the possibility of receiving a large reward induced them to feel more confident.

The overall level of accuracy was constant in both conditions and during the whole experiment, thus this increase in confidence cannot be accounted for by a parallel increase in accuracy. Moreover, subjects were accurate in approximately 70% of the cases both in the experimental and in the control condition, with no appreciable changes in confidence for 1 and for 5 €, but with a significant increase in confidence for 15 € in reward trials.

In our analysis we considered actual accuracy as a proxy of the event probability of correctly performing the experimental task. Measuring over- and underconfidence we found a significant difference between results with the lower rewards and results obtained with the highest one (1 vs. 15 €, signed-rank test $z = -2.267, P = 0.0234$; 1 vs. 5 €, signed-rank test $z = -0.937, P = 0.35$; 5 vs. 15 €, signed-rank test $z = -2.68, P = 0.0073$).

Figure 4 shows over- and underconfidence for the three different reward levels. Overconfidence (CONF – % correct answers >0) appeared only with the highest reward (15 €), whereas underconfidence (CONF – % correct answers <0) was found for the two lowest rewards. Underconfidence was found also in the control group, in line with the results about confidence judgements in perceptual tasks reported in the literature (for a review, see Baranski & Petrusic, 1994).

Regarding the control group, we found an accuracy level (73% of correct responses) in line with the average difficulty of the task, and we also found underconfidence (CONF – % correct answers = −0.04), as predicted by theories of underconfidence in perceptual tasks.

These findings confirm our prediction that there exists a desirability bias, which overcomes accuracy and induces people to rely on a possible reward more than on actual accuracy.

Moreover, these results show the effect of relative reward on confidence judgements.

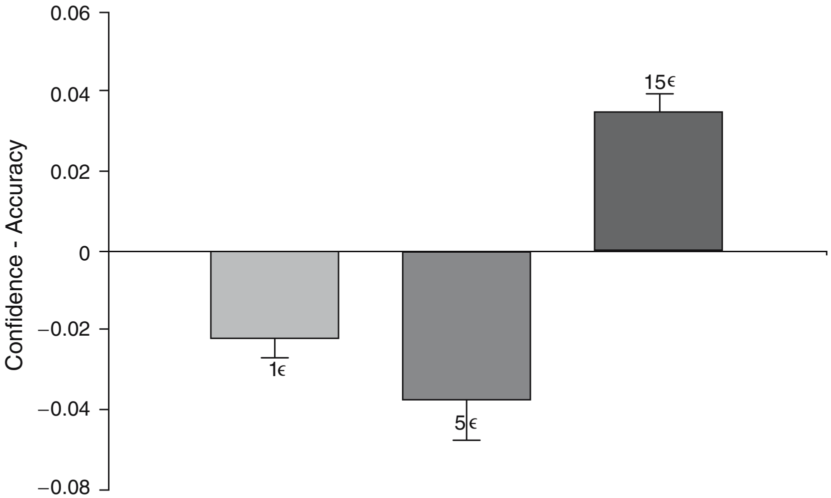

Fig. 4 Confidence (CONF/4) – Accuracy (%correct) for the three reward amounts. Results show underconfidence for 1 and 5 € and overconfidence for 15 €

3 Study

In this study, we investigated whether reward effect and desirability bias are present for other intervals of uncertainty. We tested subjects for three different difficulty levels (*Easy*, *Intermediate* and *Difficult*) and two rewards (2 and 10 €). We reduced the number of rewards, since in Study 1 we did not find any significant differences between the two lower rewards.

3.1 Participants

One hundred twenty-three undergraduate students (58 female and 65 male) from the University of Siena were recruited and randomly assigned to one of six groups. Three groups of subjects participated in reward conditions (25 for each condition, *Easy $*, *Intermediate $* and *Difficult $*) and the others three groups took part in no-reward conditions (16 control subjects for each condition, *Easy no$*, *Intermediate no$* and *Difficult no$*). All subjects were naïve with respect to the nature and aims of the experiment. Mean age of participants was 21.90 years (s.d. = 2.0096).

3.2 Stimuli

As in Study 1, we used a visual motion discrimination task (see Fig. 5). Three different difficulty levels were set in order to obtain three different levels of accuracy (in terms of percentage of correct responses). For subjects in the *Easy* condition we expected average accuracy to be higher than 0.75, for the *Intermediate* difficulty we expected results in the interval between 0.5 and 0.75, and for the *Difficult* condition we expected average accuracy to be lower than 0.50.

We assessed these conditions in a previous pilot study, where we singled out three levels of accuracy (i.e. percentage of correct answers) by manipulating the ratio between the velocity of the jiggling movement of the dots in the background, and that of the linear movement toward one out of four directions that participants had to identify (as it is in Study 1). The dots moved jiggling in a random manner toward one of four directions (right, left, up or down), whereas a fixed percentage of them moved coherently toward only one direction. We introduced two more directions (up and down) in order to increase the difficulty level.

Participants in the experiment had to identify the "coherent direction", individuating the coherently moving dots out of the background movement. The stimulus difficulty was set at the beginning of the experiment for each condition and then it was kept constant throughout the experiment. Each stimulus was presented for 2 s. The stimulus direction was randomized and controlled by the computer program, thus each of the four directions were equally probable and their single probability of occurrence was 25%.

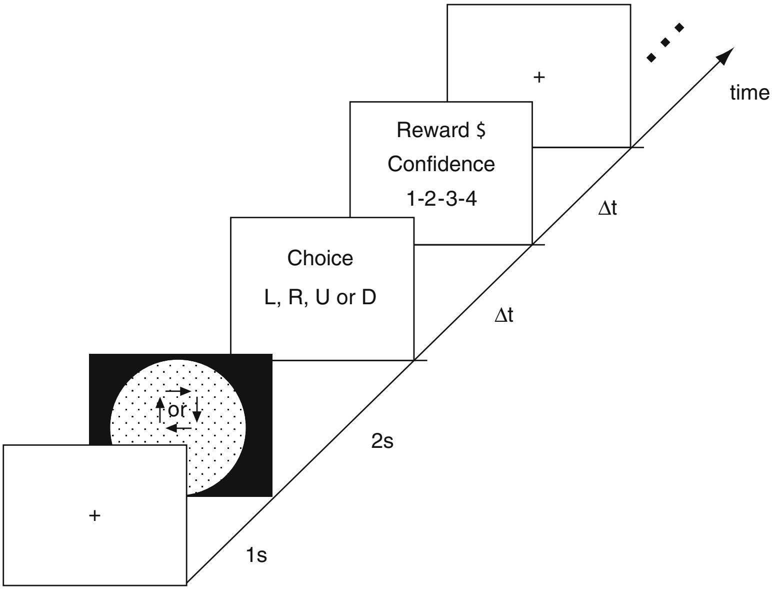

Fig. 5 Time course of the experiment in Study 2: fixation point (1 s), stimulus presentation (2 s), direction selection (up, down, left or right; self-paced), reward information (2 or 10 €), and confidence scale (self-paced)

3.3 Procedure

The sequence of events and the time course of the study was the same as in Study 1, thus [fixation point → stimulus → direction choice → possible reward → confidence judgement] → [fixation point → (...)].

The subjects were tested during six separate sessions and each and every subject participated in only one session.

Each trial began with a 1 s fixation point followed by the stimulus (2 s). After the stimulus presentation ended, participants saw on the screen the following question: "What was the direction of the dots? Left – Right – Up – Down". They responded by pressing the corresponding arrow on the PC keyboard (self-paced choice). Then, the screen was cleared and the reward and confidence scales appeared.

In the *reward conditions*, 2 and 10 € were equally probable, and subjects were instructed that the computer program randomly paired rewards with stimuli. The difficulty level for the stimulus recognition was always the same during the experiment and no correlation existed among reward amounts and stimuli. In order to avoid learning effects, no feedback was provided to participants. Subjects performed 72 trials (8 training and 64 experimental trials).

At the end of the session the subjects in the three *reward conditions* ($) completed a questionnaire and then they were paid accordingly to their performance in one trial randomly drawn by the computer out of the 64 trials (the eight training trials were excluded). If their response in the drawn trial was correct they were paid according to the reward showed during that trial, otherwise they received only the participation fee (3 €).

In three separate sessions, three control groups performed an identical task (8 training trials and 64 experimental trials) with the same three difficulty levels, except for the reward, which was neither mentioned nor displayed. These subjects were simply requested to individuate the direction and to assess their confidence on the 4-point scale (described above). At the end of the experiment they were asked to complete a questionnaire. Participants in the control groups received a show-up fee of 3 €. Each session lasted approximately 30 min.

3.4 Questionnaires

In this study we introduced a questionnaire to investigate the perceived difficulty and the determinants of subjects' confidence both in the reward and in the no-reward condition. In the former case the questionnaire was presented before subjects were informed about their winning, in order to avoid any kind of motivational or affective effects.

The questionnaire consisted of three questions regarding: difficulty (Question one: "According to you, the task was: (1) very easy; (2) fairly easy; (3) fairly difficult; (4) very difficult; (5) impossible"), accuracy (Question two: "According to you, what was the percentage of correct responses you gave?", subjects responded by circling the chosen percentage on a ten-point scale); and confidence (Question three: "According to you, which of these elements determined your confidence judgement?" – "(a) the perception of the stimulus; (b) the time required to make your choice; (c) the amount of the possible win; d. the perception of the stimulus and the amount of the possible win"). The questionnaire for the control groups was identical except for Question three, where any reference to reward was excluded (Question three: "According to you, which of these elements determined your confidence judgement?" – "(a) the perception of the stimulus; (b) the time required to make your choice").

The rationale for introducing questionnaires was the need to compare the 'trial by trial' evaluation (significantly and unequivocally affected by the displayed rewards), with the global estimate of difficulty, perceived accuracy and confidence. In other words, we were interested in assessing whether the participants, at least at the end of the task, were aware of the desirability bias. Moreover, questionnaires provided a subjective evaluation of the objective accuracy participants achieved.

3.5 Results

This study confirmed the results of Study 1, showing that the confidence level for a correct response varied with different reward levels (2 or 10 €) (data from reward condition; repeated measures ANOVA, $F(1, 74) = 50.15$, $P < 0.00001$).

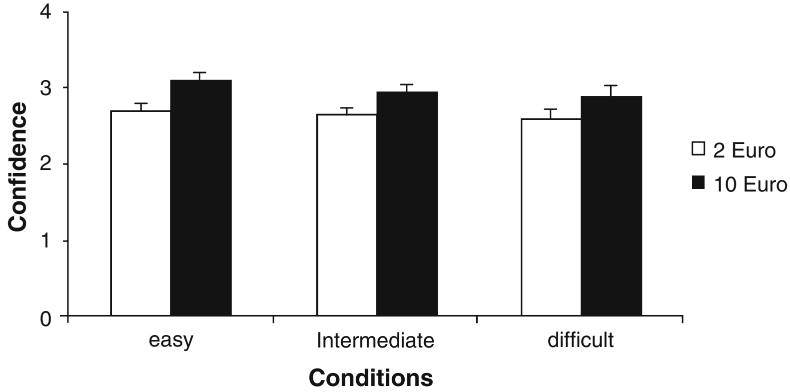

Fig. 6 Mean confidence level (+standard errors) as a function of the possible reward (2 or 10 €) for each condition (easy, intermediate, and difficult). Confidence increased as reward increased in each of the three conditions

Figure 6 shows how confidence was significantly higher for 10 € with respect to 2 €, in each condition (*Easy*, Wilcoxon signed-rank test z = 3.396, $P = 0.0007$; *Intermediate*, signed-rank test z = 4.315, $P < 0.0001$; and *Difficult*, signed-rank test z = 4.112, $P < 0.0001$).

The mean accuracy level was 0.84 (SD = 0.055), 0.61 (SD = 0.068) and 0.38 (SD = 0.073), for easy, intermediate and difficult condition, respectively. Thus, accuracy was significantly different for the three different difficulty levels (Kruskal–Wallis $Chi^2(2)$ = 63.563, $P = 0.0001$, data from reward condition; and Kruskal–Wallis $Chi^2(2)$ = 24.239, $P = 0.0001$, data from no-reward condition). However, there was no significant difference between accuracy in reward and control conditions for each level of difficulty of the task (two-sample Wilcoxon rank-sum test z = 0.866, P = 0.38, for easy; z = −0.414, P = 0.68, for intermediate; and z = 0.161, P = 0.87, for difficult). We found again a significant inverse correlation between RT and confidence level (Fig. 7 data from all conditions).

The Regression analyses using the data from the reward conditions (Order Probit model, Table 2) show that for the easy and difficult conditions the confidence was a function of the accuracy, reward level, and reaction time (inversely related). The results from the intermediate condition show that confidence judgements depended only on reward level and on reaction time (inversely related).

Thus in this condition we found, as in Study 1, that rewards by-passed the effect of the actual accuracy and biased subjects' confidence level. Results from the control conditions confirm that without the presence of rewards the determinants of confidence judgements are always accuracy and reaction time (inversely related). Furthermore, this study demonstrated that the *desirability bias* remains stable for different levels of difficulty of the task.

Figure 8 shows the pattern of over and under-confidence for different levels of difficulty of the task. In the Easy condition we found underconfidence (CONF – % correct answers <0) for both levels of rewards (2 and 10 €); in the difficult

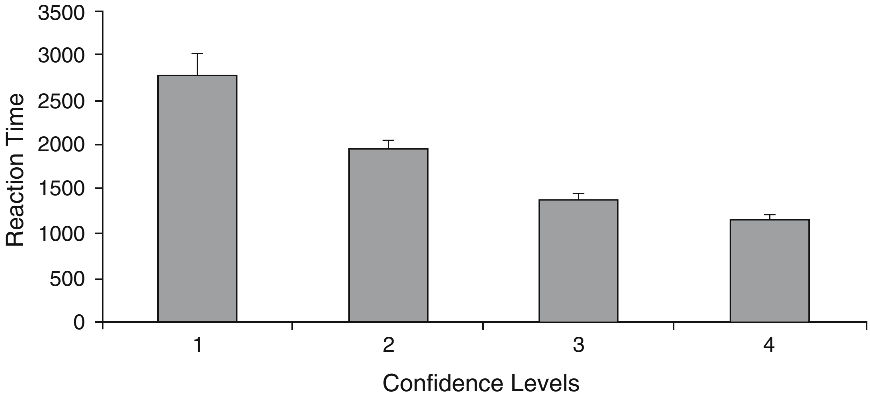

Fig. 7 Mean reaction time (RT; data collapsed over reward and no-reward conditions) for each level of confidence judgement of correct answer (with 1 – "Not sure at all" and 4 – "Really sure"). RT was inversely correlated with Confidence level, as in Study 1

Table 2 Study 2 regression analysis. Confidence levels as a function of Accuracy, reward level (only in reward condition) and reaction time. Regression analysis (Order Probit). The dependent variable Confidence takes values of 1–4. A = accuracy, equal to 1 if correct and 0 if incorrect; $ = reward level (2 or 10); *RT* reaction time

Regression analysis order probit: dependent variable is "*CONFIDENCE*"

a. *Data from experimental sessions with rewards for the three levels of difficulty (easy, intermediate, and difficult)*

	Easy				Intermediate				Difficult			
Variable	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|
A	0.2300	0.0646	3.56	0	0.0482	0.0501	0.96	0.336	0.1350	0.0552	2.44	0.015
$	0.0435	0.0069	6.3	0	0.0603	0.0063	9.62	0	0.0569	0.0068	8.33	0
RT	−0.0002	0.00002	−8.5	0	−0.0001	0.00001	−9.4	0	−0.0001	0.00002	−7.0	0

b. *Data from control sessions with no reward for the three levels of difficulty (easy, intermediate, and difficult)*

	Easy				Intermediate				Difficult			
Variable	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|	Coeff.	Std error	Z	P > |z|
A	0.6687	0.0778	8.59	0	0.3388	0.0690	4.91	0	0.2749	0.0718	3.83	0
RT	−0.0004	0.00003	−10.7	0	−0.0001	0.00002	−4.9	0	−0.0002	0.00003	−8.7	0

condition the result was inverted, thus subjects were always overconfident (CONF – % correct answers >0); whereas in the intermediate condition we found overconfidence when the reward was 10 €, and approximately calibrated judgements for the cases in which the reward was 2 €. Note that the average accuracy in the intermediate condition was slightly lower (61%) compared to the average accuracy observed in Study 1 (71%). We observe (Fig. 9) underconfidence for the lowest reward (2 €) and overconfidence for the highest reward (10 €) for level of accuracy equal to 65%.

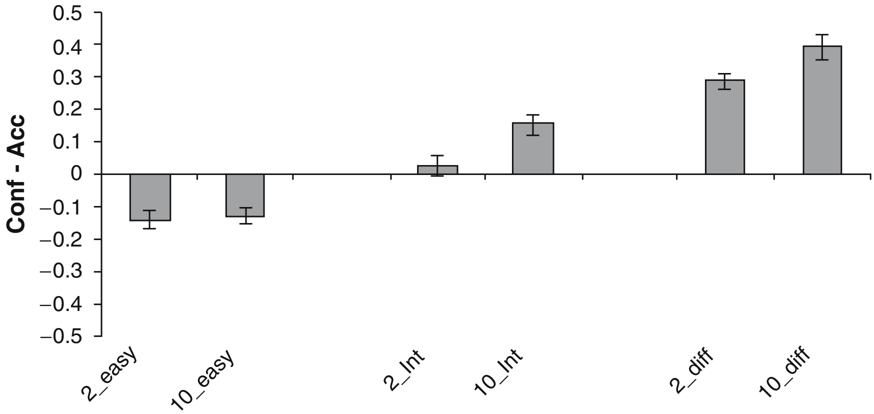

Fig. 8 Confidence (CONF/4) – Accuracy (%correct) for each difficulty level (easy, intermediate, and difficult) and for the two reward amounts (2 and 10 €)

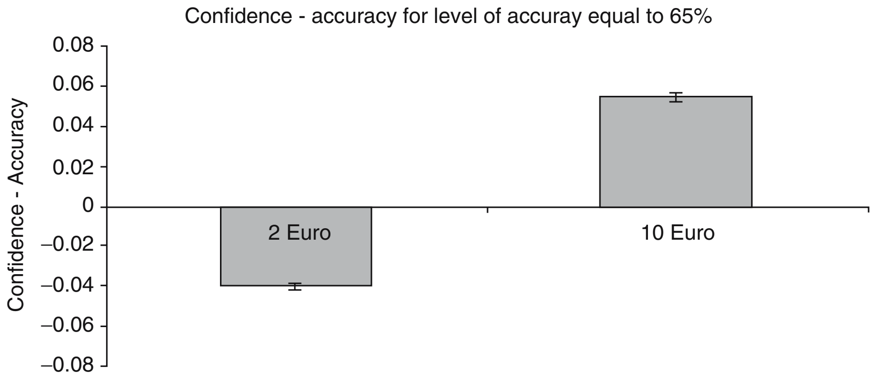

Fig. 9 Confidence (CONF/4) – Accuracy (%correct) for level of accuracy equal to 65%. As in study 1, we found overconfidence for the highest reward (10 €) and underconfidence for the lowest reward (2 €)

The analysis of the questionnaires allowed us to compare the trial-by-trial performances of subjects with their overall evaluation of their own choices. We were interested in checking whether subjects were aware or not of the role of reward and of the stimulus difficulty. On average, subjects considered the task quite difficult (mean difficulty evaluation = 3.43, SD = 0.94), with a significant difference in the relative frequencies of the responses (ranging from 1-very easy to 5-very difficult) for the three conditions (Question 1, $Chi^2(8) = 10.60$, $P = 0.22$). Figure 10 and b shows the discrepancy between Experimental Accuracy (the percentage of correct responses people gave) and the Reported Accuracy (the percentage of correct responses they thought they gave), in reward (Wilcoxon signed-rank test, $z = 7.526$, $P < 0.0001$) and no-reward conditions (Wilcoxon signed-rank test, $z = 5.56$, $P < 0.0001$) (Question 2). The difference between Experimental and

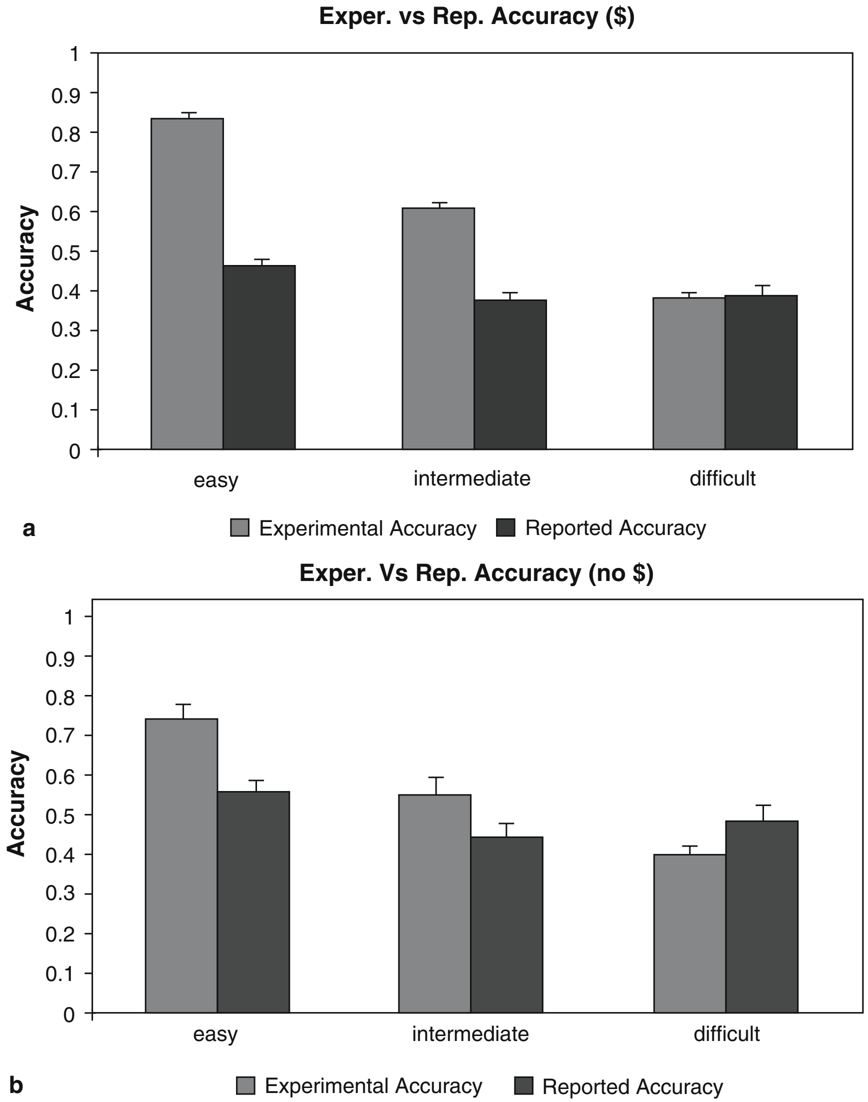

Fig. 10 **a** Comparison between Experimental Accuracy (percentage of correct responses during the task) and Reported Accuracy (estimated percentage of correct responses at the end of the task) shows that people underestimated their performances in all conditions. **b** Comparison between Experimental Accuracy (percentage of correct responses during the task) and Reported Accuracy (estimated percentage of correct responses at the end of the task) shows that people also underestimated their performances in the control condition, but this underestimation is absent in the difficult condition, where we instead found overestimation

Reported accuracy could be explained by the absence of immediate reward in the Reported Accuracy, so that people underestimated their performances, as usually happens in perceptual tasks.

The relative frequency of different reasons for confidence (Question 3: a, b, c, d) strictly depended on the difficulty level (difficult, intermediate, easy; $Chi^2(6) = 9.21$, $P = 0.16$). Almost all subjects (88%) in the Easy condition attributed their confidence to the stimulus perception, but this percentage decreases as uncertainty increased (66% intermediate, and 46% difficult). Moreover, around 30% of the subjects in the Difficult condition attributed their confidence to both the perception and the reward. This (ex-post) awareness did not prevent them from being biased by reward, as showed in Fig. 6. On the contrary, the control groups in all the three conditions attributed their confidence mainly to the stimulus perception (50% easy, 57% intermediate, 47% difficult).

4 Summary and Conclusions

People are often inaccurate in predicting their performances or their rates of success in many different domains, and many different explanations have been put forward. We suggest a general mechanism, which could work in a wide variety of domains and situations. Our findings indicate that people become relatively more confident about the occurrence of events associated with high rewards, compared with neutral events. These findings are in line with the theory of anticipatory representations by Miceli & Castelfranchi (2002), who proposed a theoretical account of expectations as a class of goal-driven anticipations.

We assume that the desirability of an outcome directly affects confidence in the occurrence of that outcome, inducing people to be more confident in it, when compared with a neutral or negative result. This assumption has been experimentally tested, and the results confirmed our hypothesis. Although the reward was merely possible, participants showed significant increases in their average confidence when a higher reward was presented. The correlation between reward and confidence was not linked to any appreciable change in accuracy, so we can reasonably conclude that the only factor modifying individuals' confidence in their choices was the reward. This means that people were not more accurate or faster in responding to the stimulus, they were just more confident in their performance when the possible reward was higher, compared to trials where the reward was lower.

Other studies (Bar-Hillel & Budescu, 1995; Irwin, 1953) tried to demonstrate the effect of a rewarding outcome on confidence levels, but motivation was not isolated from other variables, such as accuracy, so that they failed to detect any relationship between confidence and a desirable result.

By the contrary, our findings support the general hypothesis that the presence and the amount of a desirable outcome can affect people's confidence in their predictions. The pattern of confidence changes becomes especially striking when it turns into overconfidence for the highest reward in Study 1, and in the intermediate condition of Study 2 (when accuracy equal 65%). We assume that when the actual probability of the event to be predicted is extremely low or extremely high, the motivational aspects are less important in determining people's confidence judgements. Instead, when uncertainty is higher than chance but lower than certainty, judgements are desirability driven. This may happen because in this range participants were more sensitive to external (such as reward) or internal motivational cues that might drive their judgements. Considering the results of the questionnaire at the end of Study 2, we suggest that this phenomenon works at an unconscious level. Indeed, subjects indicated the perception of the stimulus as the main determinant of their confidence judgements, whereas they did not recognize the actual effect of reward.

The *desirability bias* affects people's confidence, inducing them to be more confident in the occurrence of a positive outcome, compared with a neutral one. Similar results have been reported in the psychological literature regarding "positive illusions" (Taylor & Brown, 1988), i.e. unrealistic positive beliefs about the self and one's own possibility of success and well-being. These illusions seem to be quite pervasive in human life. However, their causes are not entirely clear and the main question is whether they exert a positive or a negative influence on people's choices, behaviours and lives.

We predict that the desirability bias is a general phenomenon which could play a role in explaining optimistic overconfidence in predictions. People overestimate their possibility of achieving positive results because the "desirability bias" affects their confidence, causing them to believe that the desired result is more easily achievable. In other words, people do not simply expect events, but they actively desire positive outcomes, thus feeling more confident in the possibility of achieving the desired result.

This can be true also when the reward is not materially but psychologically relevant, such as self-esteem, social approval, and even cognitive dissonance reduction or avoidance (Blanton, Pelham, DeHart, & Carvallo, 2001; Festinger, 1957). This result is not trivial, and it could help preventing lots of mistakes due to overwhelming confidence in one's own capabilities and possibilities of success.

Acknowledgements We thank Cristiano Castelfranchi, Maria Miceli and Karen Reilly for helpful comments and suggestions. This work was funded by the PAR project 2003 "Anticipatory minds" – University of Siena (to FG); Action Concertée Incitative, Systemes Complexes from CNRS (to GC and AS); and Human Frontier Science Program (RGP 56/2005; to GC and AS). We are extremely grateful to Francesco Lomagistro for his assistance in recruiting subjects and running the experiments.

References

Ayton, P., & McClelland, A. G. R. (1997). How real is overconfidence? *Journal of Behavioral Decision Making, 10*, 279–285

Baranski, J. V., & Petrusic, W. M. (1994). The calibration of confidence in perceptual judgements. *Perception & Psychophysics, 55*, 412–428

Bar-Hillel, M. (1980). The base rate fallacy in probability judgements. *Acta Psychologica, 44*, 211–233

Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics, 54*, 75–81

Blanton, H., Pelham, B. W., DeHart, T., & Carvallo, M. (2001). Overconfidence as dissonance reduction. *Journal of Experimental Social Psychology, 37*, 373–385

Brenner, L. A., Koheler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgements: A critical examination. *Organizational Behavior and Human Decision Processes, 65*, 212–219

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*, 4745–4765

Camerer, C., & Lovallo, D. (1999). Overconfidence and excess entry: An experimental approach. *The American Economic Review, 89*, 306–318

Celebrini, S., & Newsome, W. T. (1994). Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. *Journal of Neuroscience, 14*, 4109–4124

Festinger, L. A. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row & Peterson

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

Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology, 52*, 464–473

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435

Harvey, N. (1997). Confidence in judgment. *Trends in Cognitive Science, 1*, 78–82

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226–246

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswickian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*, 344–366

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review, 103*, 582–591

Keren, G. (1988). On the ability of monitoring non-veridical perceptions and uncertain knowledge: Some calibration studies. *Acta Psychologica, 77*, 217–273

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*, 216–247

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

Kunda, Z. (1990). The case for motivated inference. *Psychological Bulletin, 108*, 480–498

Mahajan, J. (1992). The overconfidence effect in marketing management predictions. *Journal of Marketing Research, 29*, 329–342

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making, 17*, 281–295

Miceli, M., & Castelfranchi, C. (2002). The mind and the future: The (negative) power of expectations. *Theory & Psychology, 12*, 335–366

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143–148

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193–210

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*, 806–820