

The neural basis of belief updating and rational decision making

Anja Achtziger,¹ Carlos Alós-Ferrer,² Sabine Hügelschäfer,³ and Marco Steinhauser³

¹Chair of Social and Economic Psychology, Zeppelin University, 88045 Friedrichshafen, Germany, ²Department of Economics and ³Department of Psychology, University of Konstanz, 78457 Konstanz, Germany

Rational decision making under uncertainty requires forming beliefs that integrate prior and new information through Bayes' rule. Human decision makers typically deviate from Bayesian updating by either overweighting the prior (conservatism) or overweighting new information (e.g. the representativeness heuristic). We investigated these deviations through measurements of electrocortical activity in the human brain during incentivized probability updating tasks and found evidence of extremely early commitment to boundedly rational heuristics. Participants who overweight new information display a lower sensibility to conflict detection, captured by an event related potential (the N2) observed around 260 ms after the presentation of new information. Conservative decision makers (who overweight prior probabilities) make up their mind before new information is presented, as indicated by the lateralized readiness potential in the brain. That is, they do not inhibit the processing of new information but rather immediately rely on the prior for making a decision.

Keywords: Bayesian updating; conservatism; representativeness heuristic; LRP; N2

Human decision makers, from physicians and judges to firm managers and policy makers, are confronted with overwhelming amounts of information on uncertain outcomes and have to rely on predictors of only partial reliability. Reaching an optimal decision requires an appropriate integration of all available information. From a normative point of view, rational decision makers should optimize their objective functions based on beliefs updated through Bayes' rule (Bayesian updating), which captures the integration of new information with previous beliefs; this is, for instance, the classical paradigm in economics (Mas Colell *et al.*, 2005). These previous beliefs, also called priors, concern the likelihood of uncertain events. Examples range from the probability of getting infected with a certain disease to the base rate in a judgment problem. In most cases, priors are predictions people hold about probabilities of events because of previous knowledge. When additional information is acquired to make a decision (e.g. results of a medical test when considering whether having surgery), this further information should be taken into account to determine an updated probability of an uncertain event. This process should lead to an updating of priors to the so called posteriors (i.e. the probability of having caught a specific disease given the results of a medical test). This process of weighting the base belief with new evidence is described by Bayes' rule.

Although Bayes' rule is sometimes a good approximation of human behavior (El Gamal and Grether, 1995; Griffiths and Tenenbaum, 2006, 2011), a number of well documented systematic violations of Bayes' rule in conditional probability judgments show that human beings are not Bayesian optimizers (Kahneman and Tversky, 1972; Grether, 1980; Fiedler, 1988, 2000; Gigerenzer and Hoffrage, 1995; Fiedler *et al.*, 2000; Erev *et al.*, 2008). For instance, Ouwensloot *et al.* (1998) examined Bayesian updating in a semistatistical context and

observed that participants did not correctly apply Bayes' rule but, instead, systematically made errors in probability updating.

Failing to properly integrate information results in suboptimal behavior and can have detrimental effects in many areas, from medical and legal decision making to business or military contexts. Accordingly, deviations from Bayesian updating have received a great deal of attention in and beyond psychology, e.g. in economics (Camerer, 1987; Ganguly *et al.*, 2000). Determining the extent and origin of such deviations requires a better understanding of the underlying processes. The objective of the present study is to demonstrate how the measurement of brain potentials in the electroencephalogram (EEG) can contribute to this research program.

REPRESENTATIVENESS AND CONSERVATISM

Bayes' rule precisely balances prior probabilities with new information that is presented in a decision situation. Hence, a decision maker can make two kinds of mistakes: overweighting the prior (conservatism) and overweighting new information (base rate neglect). A classical example of base rate neglect is the representativeness heuristic (Kahneman and Tversky, 1972; Grether, 1980, 1992), which confounds the probability of an event with its similarity to a population and in which base rates are largely ignored. The lawyers engineers problem (Kahneman and Tversky, 1973) is a nice illustration of this heuristic. Participants were given stereotypical descriptions of alleged engineers or lawyers, supposedly extracted at random from a given set, and asked to guess the probability that a given one corresponded to a lawyer (or engineer). The base rate information (how many lawyers were in the set of available descriptions) was generally ignored in favor of the stereotypical information contained in the description.

The general phenomenon of base rate neglect (Fiedler, 2000; Fiedler *et al.*, 2000; Erev *et al.*, 2008) can be illustrated with the classical taxicab problem (Tversky and Kahneman, 1980). Here, participants should estimate the probability that a taxi of one of two companies (green and blue) was involved in an accident. Participants are told how many taxis are green (85%) and blue (15%) in this city. Moreover, they are told that a witness has identified the color of the taxi (e.g. blue) and that the probability by which this witness is able to correctly identify one of the two colors is 80% (i.e. he fails in 20% of all cases). In this

The authors thank Sander Nieuwenhuis and an anonymous referee for helpful comments. They gratefully acknowledge the financial support from the Center for Psychoeconomics at the University of Konstanz, funded within the German 'Exzellenzinitiative'. A.A. also acknowledges financial support from the German Research Foundation (DFG) under grant AC 183/2 2 (part of the Research Unit FOR 582, Limits of Intentionality).

Authors in alphabetical order. Correspondence should be addressed to Carlos Alós Ferrer, Department of Economics, Box 150, University of Konstanz, 78457 Konstanz, Germany. E mail: carlos.alos.ferrer@uni.konstanz.de

study, decision makers overweighted new information by indicating that the probability that the taxi involved in the accident was actually blue when the witness reported that color ranged from 50 to 80% while, actually, the updated probability is around 41%.

Existing explanations for these and other heuristics in probability judgments suggest that they correspond to rather automatic or impulsive processes as understood in psychology (Strack and Deutsch, 2004), i.e. being activated quickly, unconsciously and effortlessly. In contrast, processes leading to behavior aligned with Bayesian updating might be rather controlled or reflective. This view is of fundamental importance for the analysis of rational decision making, since it implies that certain decision mistakes might be associated with extremely rapid brain responses and hence be very difficult to control or train away.

To clarify the determinants of base rate neglect, we started from the idea that Bayesian updating and the representativeness heuristic correspond to different, potentially conflicting processes and hypothesized that reliance on the heuristic is associated with a low sensitivity to conflict detection at the individual level. Our reasoning was that conflict detection enables controlled processes to suppress the representativeness heuristic, and hence, subjects with lower sensitivity to conflict should be more prone to respond according to the heuristic, as this would be an automatic default. Since conflict detection occurs very early in decision making (i.e. before behavioral data can be collected), we relied on the measurement of brain potentials in the EEG, a measure of electrocortical activity that has frequently been used to investigate the temporal dynamics of decision making (Holroyd *et al.*, 2002; Yeung and Sanfey, 2004). The decision conflict should be evident in the amplitude of the N2 component of the EEG, an event related potential (ERP) capturing a negative deflection of electrocortical activity ~200–300 ms after stimulus presentation. Its amplitude reflects the degree of response conflict and is associated with activity in the anterior cingulate cortex (ACC) (Nieuwenhuis *et al.*, 2003; Folstein and Van Petten, 2008). A conflict effect in the N2 amplitude cannot reflect a response conflict between the representativeness heuristic and the ultimate outcome of Bayesian updating, because the latter process presumably takes longer than 300 ms to generate a response. Rather, it could represent a conflict between the representativeness heuristic and an inhibition process that suppresses automatically generated responses to allow the slower Bayesian updating to become effective (this type of processing architecture in decision tasks is discussed in Ridderinkhof, 2002). Such an N2 conflict effect would resemble that obtained in go/no go tasks (Nieuwenhuis *et al.*, 2003). Interindividual differences in N2 amplitude have been suggested to reflect interindividual differences in the sensitivity toward response conflict, with larger N2 amplitudes being associated with a higher sensitivity (van Boxtel *et al.*, 2001; Amodio *et al.*, 2008).

The opposite of base rate neglect is conservatism (Edwards, 1968), which amounts to overweighting the prior, ignoring or undervaluing new information. Conservative decision makers base their decisions mainly on base rates (Fischhoff and Beyth Marom, 1983; El Gamal and Grether, 1995). For example, a conservative decision maker would estimate the probability of being infected by a certain disease on the basis of the disease's known base rate of infection only, without conditioning on whether current behavior puts him or her at risk. A conservative investor might decide to invest in a particular firm or not depending on past performance only, without correcting for current market events. From a psychological perspective, conservatism in decision making is related to an attachment to past practices, beliefs (see Klayman, 1995), anchors (Mussweiler and Strack, 1999) or commitments, even when they are shown to be erroneous or detrimental.

Until today, no single explanation has been agreed upon for this phenomenon (Wallsten, 1972), which is consistently observed in

decision making studies under uncertainty. Some authors attributed conservatism to failures in information aggregation or retrieval (Edwards, 1968; Dougherty *et al.*, 1999), while others associate it with a basic, cost effective property of intuitive judgment (Kahneman, 2003), since it relies on few environmental cues. In the present study, we aimed to rule out explanations based on faulty information aggregation or retrieval in favor of the hypothesis that conservatism corresponds to a simple base rate only heuristic (Gigerenzer and Hoffrage, 1995), which disregards any additional evidence. We tested this hypothesis by analyzing the lateralized readiness potential (LRP), a waveform indicating lateralization of ERP activity over motor cortices, which reflects the central activation of a unilateral response (Eimer, 1998). A stronger amplitude of the LRP in the right hemisphere indicates an increased motor preparedness to choose the left option and vice versa. When measured prior to a stimulus, this preparedness is typically assumed to reflect a bias toward the one or the other option rather than a completed decision (Leuthold *et al.*, 1996; Steinhilber *et al.*, 2009). We hypothesized that relying on base rate information should elicit a stronger LRP, indicating which choice people are prone to, even before new information is presented. The alternative hypotheses on conservatism mentioned above would predict no differences in LRP amplitudes before presentation of the sample, because they refer to processes developing after the arrival of additional evidence.

METHODS

Participants

Twenty five participants (13 males and 12 females) with normal or corrected to normal vision, ranging in age from 19 to 34 years ($M = 21.8$, $s.d. = 2.94$), were recruited from the student community at the University of Konstanz (Germany), excluding students majoring in economics. Participants were compensated with 5 Euros plus a monetary bonus that depended upon the outcomes of the computer task. The study was conducted according to institutional guidelines, and all participants signed an informed consent document before the start of the experiment.

Procedure

Participants were tested individually in a soundproof experimental chamber and seated in front of a computer. After application of the electrodes, each participant was asked to read through the instructions explaining the experimental setup (i.e. the decision task). These instructions described the rules of the decision task in detail. The experimenter checked that the central aspects had been comprehended and clarified any misconceptions that the participant had with the rules or mechanisms. In addition, participants were instructed to move as little as possible during the computer task, to keep their fingers above the corresponding keys of the keyboard and to maintain their gaze focused at the fixation square in the center of the screen.

Decision task

The decision task is a modification of tasks used in Grether (1980, 1992), designed to test for the representativeness heuristic in an abstract, controlled framework employing minimal stimuli (hence, especially appropriate for EEG research). By construction, the task is equally well suited to test for conservative behavior. There were two urns presented on a computer screen. Urn A (left) contained three blue balls and one green, while urn B (right) contained two balls of each color. Both urns were always displayed on the same location (urn A on the left hand side of the screen and urn B on the right hand side), and in the same manner, with the top balls in blue and the bottom one(s) in green. Colors were counterbalanced (i.e. for half of the participants,

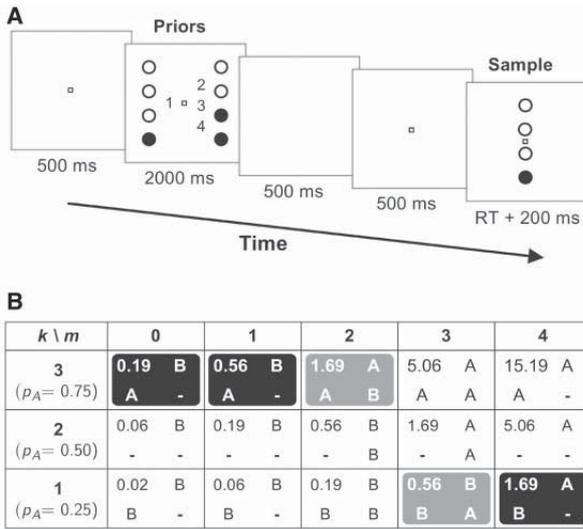


Fig. 1 Experimental design. **(A)** Sequence of events during a single trial: fixation square (500 ms), prior probabilities (2000 ms), blank screen (500 ms), fixation square (500 ms) and sample of drawn balls. The latter vanished 200 ms after the participant's response, followed by an intertrial interval (blank screen, 1500 ms). Balls shown here as black and white were green and blue in the screen, respectively. **(B)** k determines the prior for urn A (p_A), and m is the observed number of blue balls. For each (k, m) combination, the upper left hand entry gives the posterior odds for urn A, and the remaining three entries are the prescriptions for Bayesian updating (upper right hand), conservativeness (lower left hand) and the representativeness heuristic (lower right hand). Light and dark shadowed entries indicate where Bayesian updating conflicts with the representativeness heuristic and conservativeness, respectively.

urn A contained three green balls and one blue), but we will ignore this in our description to avoid confusion. In each round (Figure 1A), urn A was selected with probability $k/4$, where k varied from 1 to 3. The participant was informed of k but not of the urn actually used. This generated randomized priors with probability $1/4$, $1/2$ and $3/4$ for urn A. From the participant's perspective, this was implemented in the following way. Each urn was assigned one, two or three numbers out of four. These numbers were presented next to the urns. Then, the computer drew a random number between 1 and 4 (which was not revealed to the participant), and the urn associated with this number was selected. Subsequently, the computer extracted a sample of four random balls with replacement from the selected urn, and the participant was informed on the number of extracted blue balls, m , from 0 to 4. The four balls appeared on the screen simultaneously, stacked vertically in the order in which they had been drawn (not sorted according to color). Ignoring the order of drawn balls, this generated 15 different possible decision situations depending on the prior (k) and the sample (m). Participants were then asked to guess which urn had been actually used, using the index finger of the left hand (respectively, right hand) to press a predetermined key for urn A (respectively, urn B). Each correct answer was rewarded with 6 Euro cents at the end of the session. There was no feedback during the experiment. The prescriptions of Bayesian updating, the representativeness heuristic and conservativeness are given in Figure 1B. Two decision rules are aligned, respectively in conflict, in a (k, m) situation if they prescribe the same option, respectively different ones; note that this refers to alignment or conflict between the responses generated by the two decision rules in a particular decision situation. The representativeness heuristic only prescribes an urn if $m = 2$ or $m = 3$ (the sample looks like one of the urns), and conservativeness only applies if $k = 1$ or $k = 3$ (else the prior is 50/50). If a decision rule does not deliver a prescription for a given (k, m) situation, we say that the situation is neutral for

this rule. Figure 1B reports the odds in favor of urn A, i.e. the quotient between the probabilities for urns A and B conditional on the observation of m blue balls, given that the prior was $k/4$. Bayesian updating prescribes to choose A or B if these odds are larger or smaller than one, respectively. The paradigm provides an inverse measure of the difficulty of each (k, m) situation, in the form of odds in favor of the most likely urn after observation of the sample, i.e. the odds for A if they are larger than 1 and their inverse otherwise. Hence, the six situations with odds (for A) 1.69 or 0.56 (i.e. 1.79 for B) have a comparable difficulty, while other situations are simpler.

Each participant completed six practice trials under supervision of the experimenter to become accustomed to the computer program. Correct decisions in these trials were not rewarded. During the first three practice trials, the temporal sequence of events was decelerated. After the experimenter established that the participant completely understood the whole procedure, the Bayesian updating experiment was started, during which the EEG was recorded. For the duration of the task, the participant was alone in the experimental chamber. There were 600 trials divided in six parts, with a break of 2 min between two parts.

When all trials were completed, the amount of money earned during the task was displayed on the screen, along with details of how many decisions had been correct. Depending on the time the participant took for his/her decisions, the experiment lasted ~ 70 min. When the experimental procedure was completed, the cap and external electrodes were removed from the participant. After the computer experiment, participants filled out a questionnaire comprising several questions about personality characteristics, skills and demographic information. Finally, the participants were thanked, paid and debriefed.

EEG procedures

EEG acquisition

Data were acquired using BioSemi Active II system (BioSemi, Amsterdam, The Netherlands, www.biosemi.com) and analyzed using Brain Electrical Source Analysis (BESA) software (BESA GmbH, Grafelfing, Germany, www.besa.de) and EEGLAB 5.03 (Delorme and Makeig, 2004). The continuous EEG was recorded using 64 Ag AgCl pin type active electrodes mounted on an elastic cap, arranged according to the 10/20 system, and from two additional electrodes placed at the right and left mastoids. Eye movements and blinks were monitored by electro-oculogram (EOG) signals from two electrodes, one placed ~ 1 cm to the left side of the left eye and another one ~ 1 cm below the left eye (for later reduction of ocular artifacts). As per BioSemi system design, the common mode sense and driven right leg electrodes were used as reference and ground electrodes. Both EEG and EOG were sampled at 256 Hz. All data were re-referenced offline to an average mastoid reference and corrected for ocular artifacts with an averaged eye movement correction algorithm implemented in BESA software.

EEG analysis

Lateralized readiness potential. To analyze conservatism, we should compare stimuli with $k = 1$ and $k = 3$, independent of the sample information m . Stimulus locked data were segmented into epochs from 3100 ms before to 200 ms after stimulus onset (presentation of the sample); the interval of 100 ms before presentation of prior probabilities was used for baseline correction. Epochs for different prior probabilities were averaged separately, producing three average waveforms per participant (corresponding to $k = 1, 2$ and 3). Epochs including an EEG voltage exceeding $\pm 120 \mu V$ were omitted from

averaging to reject trials with excessive electromyogram (EMG) or other noise transients. LRPs were evaluated in two steps following a standard double subtraction method (Eimer, 1998). First, C3 C4 difference waveforms were computed for each condition of interest ($k = 1, 3$). LRPs were then computed by subtracting the waveforms for a prior of one fourth ($k = 1$) from waveforms for a prior of three fourth ($k = 3$). Those are depicted in Figure 2A. In this way, the LRP provides a relative index of conservatism, in the sense that higher values indicate a stronger action preparation for the choice of the urn with the highest prior probability (urn B for $k = 1$ and urn A for $k = 3$). Grand averages were derived by averaging these waveforms across participants. On average, 34% of trials were excluded due to artifacts, with a majority being movement related muscular artifacts. The large number of excluded epochs was due to the epochs' length, which increased the probability that a given epoch was contaminated by a muscular artifact. To quantify the LRP in the averaged ERP waveforms for each participant, the mean amplitude during the 100 ms time interval preceding stimulus onset (presentation of the sample) was calculated. This time window was chosen because it reflects participants' left-right orientation immediately before the sample is presented and a decision is required.

N2. For the representativeness heuristic, conflict situations ($k = 3, m = 2$) and ($k = 1, m = 3$) should be compared with situations of comparable difficulty ($k = 3, m = 1$) and ($k = 1, m = 4$), which are neutral for this heuristic. For the analysis of the N2, stimulus locked data were segmented into epochs from 100 ms before to 1000 ms after stimulus onset (presentation of the sample); the prestimulus interval of 100 ms was used for baseline correction. In line with previous studies (Bartholow *et al.*, 2005), only trials with correct reactions were used for data analyses. Epochs locked to the conflict stimuli ($k = 3, m = 2$) and ($k = 1, m = 3$) and the neutral stimuli ($k = 3, m = 1$) and ($k = 1, m = 4$) were averaged separately, producing two average waveforms per participant. Epochs including an EEG or EOG voltage exceeding $\pm 120 \mu\text{V}$ were omitted from averaging, to reject trials with excessive EMG or other noise transients. The difference waveform was computed by subtracting conflict waveforms from neutral waveforms. Grand averages were derived by averaging these ERPs across participants. On an average, 8% of trials were excluded due to artifacts.

To quantify the N2 in the averaged ERP waveform for each participant, the mean amplitude in the interval 235–285 ms after stimulus onset (presentation of the sample) was calculated. This time window was chosen because previous research has found the N2 peak in this period (Nieuwenhuis *et al.*, 2003; Bartholow and Dickter, 2008) and because, in our data, the peak of the N2 occurred at 260 ms in the grand average waveform. We also checked that the results of the analysis are unchanged if one uses the interval 200–320 ms instead. In accordance with previous studies, the N2 amplitude was evaluated at channel FCz, where it is normally maximal (Nieuwenhuis *et al.*, 2003; Bartholow *et al.*, 2009). Similar effects were obtained when channel Cz was analyzed.

RESULTS AND DISCUSSION

Table 1 summarizes the error rates and the median response times for all participants across all 15 decision situations. Focus on the two situations with ($k = 2, m = 2$) and ($k = 2, m = 3$), where there is no response conflict among the postulated processes. Compared with these situations and as observed in behavioral studies since Grether (1980, 1992), error rates are higher in situations of comparable difficulty, where Bayes' rule conflicts with the representativeness heuristic [($k = 3, m = 2$) and ($k = 1, m = 3$)]; Wilcoxon signed rank test for paired samples ($z = 2.76, P = 0.006$) or with conservatism

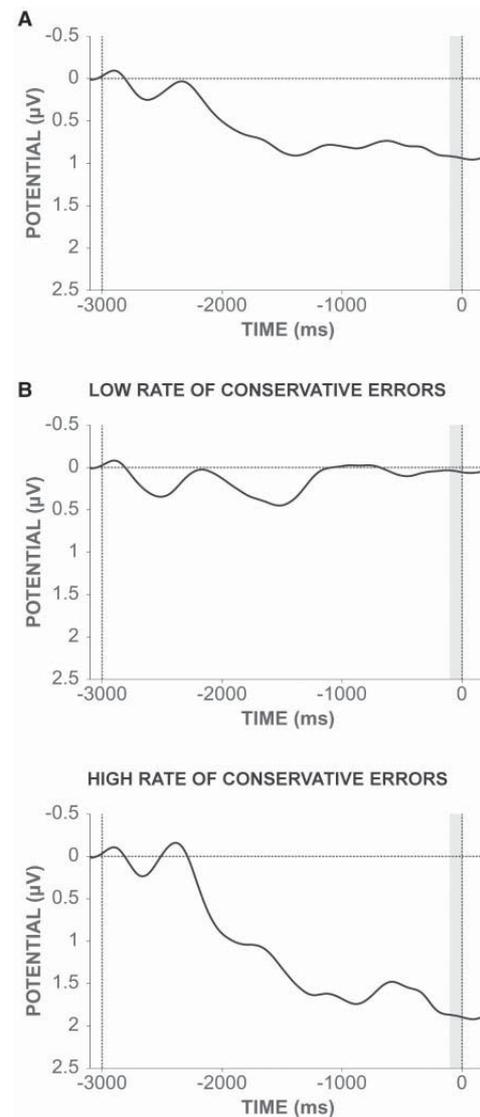


Fig. 2 The LRP. (A) Grand average stimulus locked LRP waveforms ($k = 3$ minus $k = 1$). Time zero corresponds to sample presentation; time -3000 to the presentation of prior probabilities. The more positive the value of the LRP amplitude, the stronger is the hand specific activation of the response with the higher prior probability. Light gray bar indicates time window used for analyses. (B) Grand average stimulus locked LRP waveforms for participants with low vs high rates of conservative errors (median split).

[($k = 3, m = 1$) and ($k = 1, m = 4$); $z = 2.44, P = 0.015$].¹ We can also observe that median response times are precisely highest in the four conflict situations just mentioned, as should be expected, since conflict resolution is time consuming. This intuition is confirmed by a regression analysis on response times, which we will report below.

Figure 3A shows the grand average waveforms from the frontocentral electrode FCz, depending on whether there is a conflict between Bayesian updating and representativeness or not, and the corresponding difference waveforms for all participants. Starting ~ 200 ms after the presentation of the sample, the two waveforms followed a

¹We further particularized this test following a median split of participants according to their error rates in situations where the representativeness heuristic conflicts with Bayesian updating. The difference is significant for participants with high error rates ($z = 2.98, P = 0.003$) but not significant for those with low error rates ($z = 0.25, P = 0.807$). The analogous result holds for the case of conflict with conservatism, where the median split is conducted according to error rates in situations where Bayesian updating conflicts with conservatism (high error rates, $z = 3.06, P = 0.002$; low error rates, $z = 1.08, P = 0.279$).

Table 1 Descriptive error frequencies and response times for all prior-outcome combinations over all participants

<i>k</i>	<i>m</i>				
	0	1	2	3	4
3	11.9% (101) 782	17.8% (518) 906	32.2% (1379) 949	2.5% (2072) 648	0.4% (1378) 555
2	2.4% (123) 648	4.0% (531) 742	9.5% (1113) 760	13.4% (1273) 856	2.2% (727) 638
1	1.2% (256) 662	0.8% (1166) 650	1.5% (1861) 631	23.8% (1680) 953	28.8% (737) 890

The top number is the percentage of errors, with the total number of observations in brackets. The bottom number is the median response time in milliseconds.

differentiated time course. There is a more pronounced N2 for situations in which there is a conflict between Bayes' rule and the representativeness heuristic, peaking at ~260 ms, and a more pronounced P300 (P3a and P3b components) for situations that are neutral for the representativeness heuristic. The scalp topographies of the difference waveforms included in Figure 3 show the spatial distribution of these effects.

Mean amplitudes of the difference wave for neutral vs conflict situations between 235 and 285 ms were significantly different from zero (nonparametric Wilcoxon signed rank one sample test, $M = 0.97$, $s.d. = 2.37$, $z = 2.44$, $P = 0.015$).² However, the scalp topographies of the difference wave for time periods succeeding the N2 (Figure 3A) reveal that the frontocentral conflict effect is not restricted to the N2 but spans across several components. Moreover, there is also a significant difference between conflict and neutral situations at posterior electrodes, which presumably reflects a modulation of the later P300. Mean amplitudes of the difference wave for neutral vs conflict situations between 485 and 535 ms at channel Pz were significantly different from zero (Wilcoxon signed rank one sample test, $M = 4.31$, $s.d. = 5.43$, $z = 3.54$, $P = 0.000$). However, although the frontocentral and posterior conflict effects are partially overlapping in time, they can be clearly dissociated given their different temporal onsets.

We are interested in individual differences, and hence, we will turn to a regression analysis given later. As a preliminary graphical illustration, Figure 3B shows grand average waveforms and corresponding topographies separately for participants having a low or high error rate in situations in which the representativeness heuristic conflicted with Bayesian updating (following a median split). The conflict effect in the N2 period was more pronounced for participants with low error rates than for participants with high error rates. A comparison of the difference waves reveals that this might reflect that the onset of the frontocentral conflict effect is delayed rather than absent for participants with high error rates. This interpretation is compatible with brain imaging studies (De Neys *et al.*, 2008), showing that areas involved in conflict detection (the ACC) were always activated in an

²As mentioned above, we postulate that the N2 reflects a conflict between the representativeness heuristic and an inhibition process. This implies that any condition in which the representativeness heuristic prescribes a response should lead to an increased N2, even if this response agrees with Bayesian updating. To test this assumption, we additionally analyzed N2 amplitudes in those conditions in which the representativeness heuristic and Bayesian updating were aligned [($k = 2, m = 2$) and ($k = 2, m = 3$)] and found that mean amplitudes of the difference wave for our initial neutral conditions [($k = 3, m = 1$) and ($k = 1, m = 4$)] vs these alignment conditions were different from zero ($M = 0.90$, $s.d. = 2.90$, $z = 1.91$, $P = 0.056$), while they were not significantly different from zero ($M = 0.07$, $s.d. = 2.34$, $z = 0.20$, $P = 0.840$) when comparing these conditions with our initial conflict conditions [($k = 3, m = 2$) and ($k = 1, m = 3$)]. This strengthens our interpretation of the N2. Note that although the ($k = 2, m = 2$) and ($k = 2, m = 3$) conditions involve a process conflict, the responses of Bayesian updating and representativeness are aligned in these conditions, hence they are not necessarily associated with larger error rates or longer response times.

implementation of the lawyers engineers problem (Kahneman and Tversky, 1973). Altogether, this indicates that participants with an earlier onset of the frontocentral conflict effect (represented by a larger conflict related N2), and thus, with a larger sensitivity for detecting conflict early (van Boxtel *et al.*, 2001; Amodio *et al.*, 2008), were better able to avoid the kind of errors that result from an application of the representativeness heuristic.

Concerning conservativeness, we examined the grand average LRP waveforms, i.e. C3 C4 for $k = 3$ minus C3 C4 for $k = 1$ (Figure 2A). Mean LRP amplitudes (where positive values are associated with conservatism) were significantly different from zero (nonparametric Wilcoxon signed rank one sample test, $M = 0.86$, $s.d. = 2.89$, $z = 2.03$, $P = 0.042$), indicating a bias due to conservativeness for the overall sample.

Again, we will rely on a regression analysis to test for individual differences. As a graphical illustration, Figure 2B shows the grand average LRP separately for participants having a low or high rate of conservative errors (following a median split). The amplitude of this waveform is an indicator of the participant's orientation toward the urn with the highest prior probability, before the sample (new information) was presented. It was observed that, before the sample was presented, participants with a high rate of conservative errors were more strongly oriented toward the urn with the highest prior probability than participants with a low rate of conservative errors.

To quantify our results controlling for individual heterogeneity, we ran a probit regression with random effects for participants on decision errors (Table 2). Probit models (Baltagi, 2005) are latent variable models widely used when the dependent variable is binary (in our case, committing an error or not). We coded errors as 1 and correct answers as 0, and hence (following the standard interpretation of probit models), positive regressor coefficients indicate an increase in the likelihood of error when the regressor's value increases. The random effects formulation allows controlling for individual differences when the data form a panel, as is the case when one has multiple observations for every participant. At the same time, one can include participant variables as regressors. In particular, we include participant specific average LRP and N2 amplitudes (interacted with the appropriate conflict situations, i.e. conflict with representativeness for the N2 variable and conflict with conservatism for the LRP variable). The estimated model was

$$Y_{it}^* = \text{Constant} + \text{ConfRep}_{it} + \text{AlignRep}_{it} + \text{ConfCons}_{it} + \text{N2}_i \times \text{ConfRep}_{it} + \text{LRP}_i \times \text{ConfCons}_{it} + \text{odds}_{it} + (\text{odds}_{it})^2 + t + \text{cb}_{it} + \text{Gender}_i + \text{Stat}_i + \alpha_i + \varepsilon_{it},$$

where i and t are the participant and trial (round) indices, respectively, and Y_{it}^* is the latent variable, i.e. the observed binary variable Y_{it} takes the value 1 in case of error and $Y_{it} = 1$ if and only if $Y_{it}^* > 0$. The variable ε_{it} is the trial error term, and the variable α_i is the error term capturing random effects at the participant's level. Both are assumed to be independently normally distributed. The variables ConfRep and ConfCons are dummy variables, taking the value 1 in case of conflict of Bayes' rule with representativeness and conservatism, respectively. AlignRep is a dummy variable taking the value 1 in case Bayes' rule and representativeness are aligned. The variable odds is defined as the odds for the most likely urn = 1 (so that values closer to 0 indicate harder choices). The N2 and LRP variables were measured as mean amplitudes of the difference waveforms described earlier. The dummy variables cb and Gender record color counterbalance and gender (1 for male), and the variable Stat corresponds to the participant's self assessed level of knowledge in statistics.

We found a significant positive effect of LRP amplitude in situations where conservatism conflicts with Bayesian updating; i.e. in situations

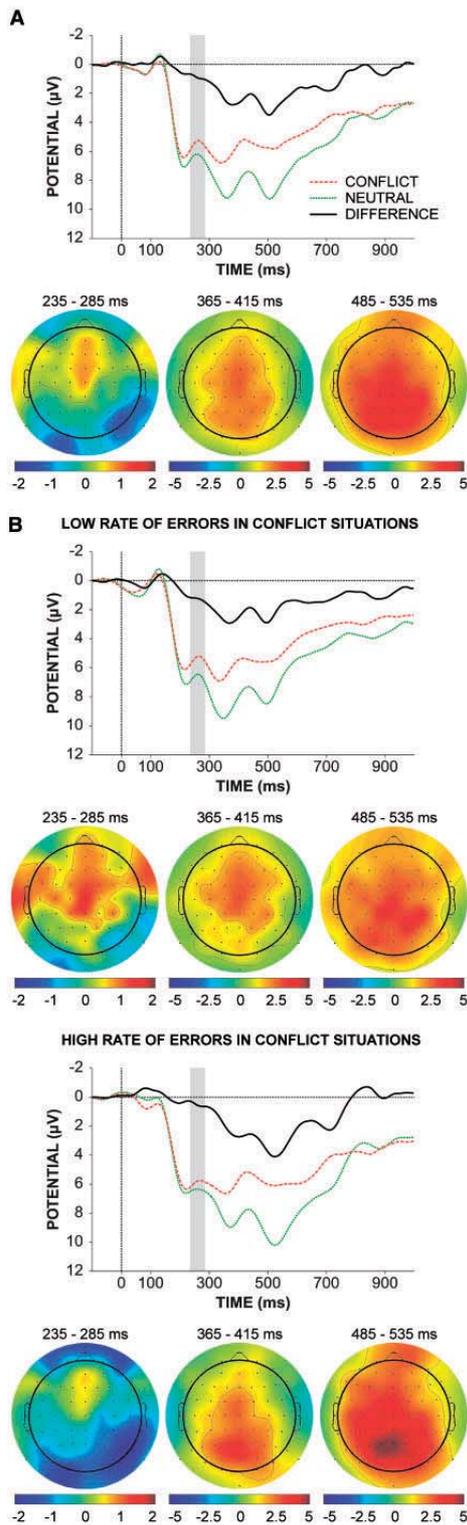


Fig. 3 The N2 component at electrode FCz. (A) Grand average N2 across all participants depending on situation type. Time zero corresponds to sample presentation. Light gray bar indicates time window used for analyses. To ensure that the increased negativity for conflict situations was indeed due to an increased N2, we considered the spatial distribution of the difference between neutral and conflict situations (for the representativeness heuristic) in this and subsequent time ranges. The picture shows that the conflict effect has the typical frontocentral distribution of an N2. Activity represents the difference between neutral and conflict situations in the indicated time ranges. (B) Grand average N2 waveforms for participants with low vs high rate of errors in situations with a conflict between Bayesian updating and the representativeness heuristic (median split).

Table 2 Random-effects probit regression on decision errors (0 = correct choice, 1 = error)

Variable	β	SE
Conflict with representativeness (1 = yes)	0.642***	0.088
Alignment with representativeness (1 = yes)	0.199**	0.078
Conflict with conservatism (1 = yes)	0.346***	0.093
N2 \times conflict with representativeness	1.098***	0.126
LRP \times conflict with conservatism	1.137***	0.160
Counterbalance	0.025	0.158
Round	0.249***	0.055
Corrected odds	0.198***	0.014
Corrected odds (squared)	0.004***	0.000
Gender (1 = male)	0.180	0.154
Knowledge in statistics	0.019	0.029
Constant	0.742***	0.197
Log likelihood	3827.981	
Wald test	1355.23***	

Number of observations = 14 915. LRP and N2 amplitudes were measured in units of $10 \mu V$ for comparability with other variables. An alternative model in which LRP and N2 amplitudes were included as separate predictor variables did not have a better fit.

* $P < 0.10$. ** $P < 0.05$. *** $P < 0.001$.

where conservatism results in an error, a larger LRP amplitude (indicating a decision toward the urn favored by the prior) is associated with an increase in the likelihood of an error. We also found a significant negative effect of N2 amplitude in situations where representativeness conflicts with Bayesian updating. That is, in situations where representativeness leads to an error, a larger N2 amplitude (indicating a higher sensitivity toward conflict detection) is associated with a decrease in the likelihood of an error.³ These results provide a test of the facts illustrated earlier while controlling for a number of factors. The regression also shows that (i) more difficult situations, as measured by the odds of the most likely urn, are associated with more errors, (ii) there is a learning effect, with errors becoming less likely over time and (iii) the presence of a conflict between representativeness and Bayesian updating has a large positive effect (increased error likelihood). All these observations are natural. For instance, in the presence of a conflict with Bayesian updating, the representativeness heuristic delivers the wrong answer, and hence, more errors are to be expected than when this conflict is absent.

We also conducted a random effects linear regression on logarithmed response times (Table 3). We found that the LRP amplitude was negatively associated with response times, whereas the N2 amplitude did not affect decision latencies. The significant effect of the LRP amplitude is as expected. Participants who were strongly oriented toward the alternative with the higher prior probability (as reflected in their LRP amplitude) already prepared a response for this alternative before the presentation of the sample evidence, which allowed for a fast response. Note that this effect on response times is independent of whether following the prior leads to a correct or an incorrect answer (hence, no interaction with a conflict dummy is necessary). Results of the regression also show that the presence of a conflict between Bayesian updating and representativeness or conservatism was associated with significantly longer response times. In contrast, alignment between Bayesian updating and representativeness significantly decreased response latencies. Decision times increased with decision difficulty (as measured by the odds of the most likely urn) and

³We ran an additional regression, including the conflict effect of the P300, measured as the mean difference between conflict and neutral situations in a time period from 485 to 535 ms at channel Pz. No significant interaction between P300 amplitude and conflict was obtained ($P = 0.896$). This further shows that the frontocentral conflict effect starting in the N2 time period and the corresponding P300 effect can be dissociated.

Table 3 Random-effects linear regression on decision response times (logarithmed)

Variable	β	SE
Conflict with representativeness (1 = yes)	0.085***	0.019
Alignment with representativeness (1 = yes)	0.088***	0.015
Conflict with conservatism (1 = yes)	0.081***	0.021
N2	0.003	0.259
LRP	0.598**	0.229
Counterbalance	0.074	0.141
Round	0.391***	0.013
Corrected odds	0.042***	0.002
Corrected odds (squared)	0.001***	0.000
Gender (1 = male)	0.291**	0.129
Knowledge in statistics	0.023	0.024
Constant	7.077***	0.154
R^2	0.243	
Wald test	3003.16***	

Number of observations = 14 915. LRP and N2 amplitudes were measured in units of $10 \mu V$ for comparability with other variables.

* $P < 0.10$. ** $P < 0.05$. *** $P < 0.001$.

decreased over the course of the experiment. Interestingly, males required a significantly longer response time than females.

In summary, our results show that reliance on heuristics in probability updating is associated with early components in the EEG, indicating an extremely quick onset of boundedly rational processes. Individual heterogeneity in the form of differences in sensitivity to detect conflicting decision rules plays an important role in situations that require Bayesian updating. Decision makers who are able to (rationally) follow Bayes' rule and suppress the automatic response of following the representativeness heuristic are more sensitive to conflict detection, with the difference to more intuitive decision makers already apparent around 260 ms after the onset of new information (the sample).

Strikingly, conservative decision makers (that is, people who strongly rely on base rate information) initiate action choice (captured by LRP amplitude) well before new information is presented. This allows us to settle a classic debate (Wallsten, 1972) on the origin of conservatism. Since conservative participants in our study had already made their decision even before new information was presented, we can rule out previous explanations attributing conservatism to a faulty aggregation of prior and sample (Edwards, 1968), fallible retrieval processes (Dougherty *et al.*, 1999) or avoidance of extreme responses (DuCharme, 1970). Our results fit alternative explanations, postulating that decision makers confronted with uncertain environments often undervalue the diagnostic impact of new evidence (for instance, the results of a medical test) and hence ignore it (Peterson and Beach, 1967; Navon, 1978; Chase *et al.*, 1998). Our study provides an example of how neuroscientific methods allow for an investigation of processes underlying decision behavior that cannot be investigated by purely behavioral methods.

REFERENCES

- Amodio, D.M., Master, S.L., Yee, C.M., Taylor, S.E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: implications for theories of self-regulation. *Psychophysiology*, 45, 11–9.
- Baltagi, B.H. (2005). *Econometric Analysis of Panel Data*, 3rd edn. Chichester, UK: John Wiley.
- Bartholow, B.D., Dickter, C.L. (2008). A response conflict account of the effects of stereotypes on racial categorization. *Social Cognition*, 26, 314–32.
- Bartholow, B.D., Pearson, M.A., Dickter, C.L., Sher, K.J., Fabiani, M., Gratton, G. (2005). Strategic control and medial frontal negativity: beyond errors and response conflict. *Psychophysiology*, 42, 33–42.
- Bartholow, B.D., Riordan, M.A., Sauls, J.S., Lust, S.A. (2009). Psychophysiological evidence of response conflict and strategic control of responses in affective priming. *Journal of Experimental Social Psychology*, 45, 655–66.

- Camerer, C.F. (1987). Do biases in probability judgment matter in markets? Experimental evidence. *American Economic Review*, 77, 981–97.
- Chase, V.M., Hertwig, R., Gigerenzer, G. (1998). Visions of rationality. *Trends in Cognitive Sciences*, 2, 206–14.
- Delorme, A., Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134, 9–21.
- De Neys, W., Vartanian, O., Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychological Science*, 19, 483–9.
- Dougherty, M.R.P., Gettys, C.F., Ogden, E.E. (1999). MINERVA-DM: a memory processes model for judgments of likelihood. *Psychological Review*, 106, 180–209.
- DuCharme, W.M. (1970). Response bias explanation of conservative human inference. *Journal of Experimental Psychology*, 85, 66–74.
- Edwards, W. (1968; reprinted 1982). Conservatism in human information processing. In: Kleinmuntz, B., editor. *Formal Representation of Human Judgment*. New York: Wiley, pp. 17–52.
- Eimer, M. (1998). The lateralized readiness potential as an on-line measure of central response activation processes. *Behavior Research Methods, Instruments, & Computers*, 30, 146–56.
- El-Gamal, M.A., Grether, D.M. (1995). Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association*, 90, 1137–45.
- Erev, I., Shimonowitch, D., Schurr, A., Hertwig, R. (2008). Base rates: how to make the intuitive mind appreciate or neglect them. In: Plessner, H., Betsch, C., Betsch, T., editors. *Intuition in Judgment and Decision Making*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 135–48.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–9.
- Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, 107, 659–76.
- Fiedler, K., Brinkmann, B., Betsch, T., Wild, B. (2000). A sampling approach to biases in conditional probability judgments: beyond base rate neglect and statistical format. *Journal of Experimental Psychology: General*, 129, 399–418.
- Fischhoff, B., Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90, 239–60.
- Folstein, J.R., Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology*, 45, 152–70.
- Ganguly, A.R., Kagel, J.H., Moser, D.V. (2000). Do asset market prices reflect traders' judgment biases? *Journal of Risk and Uncertainty*, 20, 219–45.
- Gigerenzer, G., Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychological Review*, 102, 684–704.
- Grether, D.M. (1980). Bayes rule as a descriptive model: the representativeness heuristic. *Quarterly Journal of Economics*, 95, 537–57.
- Grether, D.M. (1992). Testing Bayes rule and the representativeness heuristic: some experimental evidence. *Journal of Economic Behavior and Organization*, 17, 31–57.
- Griffiths, T.L., Tenenbaum, J.B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, 17, 767–73.
- Griffiths, T.L., Tenenbaum, J.B. (2011). Predicting the future as Bayesian inference: people combine prior knowledge with observations when estimating duration and extent. *Journal of Experimental Psychology: General*, 140, 725–43.
- Holroyd, C.B., Coles, M.G.H., Nieuwenhuis, S. (2002). Medial prefrontal cortex and error potentials. *Science*, 296, 1610–11.
- Kahneman, D. (2003). Maps of bounded rationality: psychology for behavioral economics. *American Economic Review*, 93, 1449–75.
- Kahneman, D., Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, 3, 430–54.
- Kahneman, D., Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–51.
- Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 42, 385–418.
- Leuthold, H., Sommer, W., Ulrich, R. (1996). Partial advance information and response preparation: evidence from the lateralized readiness potential. *Journal of Experimental Psychology: General*, 125, 307–323.
- Mas-Colell, A., Whinston, M.D., Green, J.R. (2005). *Microeconomic Theory*. New York: Oxford University Press.
- Mussweiler, T., Strack, F. (1999). Comparing is believing: a selective accessibility model of judgmental anchoring. In: Stroebe, W., Hewstone, M., editors. *European Review of Social Psychology*, Vol. 10. Chichester, England: Wiley, pp. 135–67.
- Navon, D. (1978). The importance of being conservative: some reflections on human Bayesian behaviour. *British Journal of Mathematical and Statistical Psychology*, 31, 33–48.
- Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., Ridderinkhof, K.R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cognitive, Affective & Behavioral Neuroscience*, 3, 17–26.
- Ouwensloot, H., Nijkamp, P., Rietveld, P. (1998). Errors in probability updating behaviour: measurement and impact analysis. *Journal of Economic Psychology*, 19, 535–63.
- Peterson, C.R., Beach, L.R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, 68, 29–46.

- Ridderinkhof, K.R. (2002). Micro- and macro-adjustments of task set: activation and suppression in conflict tasks. *Psychological Research*, 66, 312-23.
- Steinhauser, M., Hubner, R., Druey, M. (2009). Adaptive control of response preparedness in task switching. *Neuropsychologia*, 47, 1826-35.
- Strack, F., Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220-47.
- Tversky, A., Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In: Fishbein, M., editor. *Progress in Social Psychology*. Vol. 1, Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 49-72.
- van Boxtel, G.J.M., van der Molen, M.W., Jennings, J.R., Brunia, C.H.M. (2001). A psychophysiological analysis of inhibitory motor control in the stop-signal paradigm. *Biological Psychology*, 58, 229-62.
- Wallsten, T.S. (1972). Conjoint-measurement framework for the study of probabilistic information processing. *Psychological Review*, 79, 245-60.
- Yeung, N., Sanfey, A.G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, 24, 6258-64.