Individual differences in the perception of cue-outcome contingencies: a signal detection analysis.

Jeremie Jozefowiez

To cite this version:

Jeremie Jozefowiez. Individual differences in the perception of cue-outcome contingencies: a signal detection analysis.. Behavioural Processes, 2021, Behavioural Processes, 188, p.104398. 10.1016/j.beproc.2021.104398 . hal-04301682

HAL Id: hal-04301682
https://hal.univ-lille.fr/hal-04301682v1

Submitted on 22 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Version of Record: https://www.sciencedirect.com/science/article/pii/S0376635721000851
Manuscript_00f06f4493b34bf143e0fb5e220f1e5d

Running head: INDIVIDUAL DIFFERENCES IN THE PERCEPTION OF CUE-OUTCOME CONTINGENCIES

Individual Differences in the Perception of Cue-Outcome Contingencies: A Signal Detection Analysis

Jérémie Jozefowiez,

Université de Lille

Correspond with:
Dr. Jérémie Jozefowiez
Univ.Lille, CNRS UMR 9193 – SCALab – Sciences Cognitives et Affectives
F-59000 Lille
France
Email: jeremie.jozefowiez@univ-lille.fr
Phone: +33-320-041-6866
Fax: +33-320-041-6036

Submitted: December 2020
Revised: April 2021

© 2021 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

Abstract

In a signal detection theory (SDT) approach to associative learning, one assumes that, when a subject is exposed to a flow of stimuli, an association is created between the internal representations of a cue and of an outcome, allowing the representation of the cue to activate the representation of the outcome. The outcome activation is a random variable drawn from a Gaussian distribution with mean *m* (sensitivity to the contingency) and standard deviation *d* (variability in outcome activation). Depending on whether the outcome activation is above or below various decision thresholds, the participant perceives either a negative, a null, or a positive contingency between the cue and the outcome. This study presents a detailed SDT analysis of the performance of four participants on whom data in a contingency assessment task were collected almost daily during several months. Parameters from the SDT model proved relatively stable over time, except if feedback was provided to the subject. In that case, for some participants but not all, the sensitivity increased. The decision criteria were also affected. Some of these changes endured despite the discontinuation of feedback. The variability in outcome activation was not affected by the feedback.

Keywords: associative learning, contingency assessment, streamed-trial procedure, signal detection theory, single-subject design.

In a contingency assessment task, a participant is exposed to a sequence of stimuli before judging whether one of them (the cue) allows to predict another (the outcome). Contingency judgments are supposed to reflect the strength of an association that has built up between the internal representations of the cue and of the outcome (associative learning). An association is the ability of the representation of the cue to influence the activation of the representation of the outcome (hence worth, the outcome activation). The same associative process is thought to explain the development of a conditioned response to a conditioned stimulus in Pavlovian conditioning (Shanks, 1995).

An important variable affecting the strength of the cue-outcome association is the objective predictive relation (contingency) between the cue and the outcome which can be measured through the ΔP index. ΔP is the difference between the occurrence of the outcome in the presence of the cue and its occurrence in the absence of the cue. When ΔP is positive (positive contingency), the cue is a genuine predictor of the occurrence of the outcome; when ΔP is equal to 0 (null contingency), the cue does not allow to predict either the occurrence nor the non-occurrence of the outcome; when ΔP is negative (negative contingency), the cue is a genuine predictor of the non-occurrence of the outcome. Contingency judgments are a function of ΔP: participants are more likely to perceive a predictive relation between the cue and the occurrence of the outcome when ΔP is positive as opposed to null or negative; they are more likely to perceive a predictive relation between the cue and the non-occurrence of the outcome when ΔP is negative as opposed to null or positive (see Schanks, 1995, 2007 for a review).

Allan et coll. (Allan, Hannah, Crump, & Siegel, 2008; Allan, Siegel, & Tangen, 2005; Siegel, Crump, & Allan, 2009. See also Laux, Godaert, & Markman, 2005; Maia, Lefèvre, & Jozefowiez, 2018 and Perales, Catena, Schanks, & Gonzales, 2005) have proposed an analysis of contingency assessment within the framework of signal detection theory (SDT.

See Wickens, 2002 for an overview of SDT): after a subject is exposed to a sequence of stimuli, the outcome activation is a random variable drawn from a Gaussian distribution with mean M and a standard deviation of D; when it falls above a critical value C, the participant perceives a predictive relation between the cue and the occurrence of the outcome. Such an approach has two inherent advantages. First, contrary to the major theories of associative learning which are all deterministic (i.e. Mackintosh, 1973; Miller & Matzel, 1988; Pearce, 1987; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Wagner, 1981), the SDT approach assumes that the outcome activation is inherently variable and tries to quantify this variability. On the other hand, it is agnostic concerning the source of this variability: it might come from the outcome representation itself, the strength of the cue-outcome association, or the cue representation. Second, it does not assume a one-to-one mapping between outcome activation and performance. Contingency judgments are the results of three different processes encapsuled in the three sets of SDT parameters: the sensitivity of the participant to the cue-outcome contingency as expressed by the mean M of the distributions of outcome activation; the variability in the outcome activation as expressed by the standard deviation D of the distributions of outcome activation; and the decision criteria expressed by the critical value C.

SDT data analysis techniques allow experimenters to retrieve these SDT parameters from the data (see Wickens, 2002). Applied to contingency assessment, this requires a precise estimate of the probability for the participant to identify a positive contingency between the cue and the outcome when presented with a specific value of ΔP. It would take hours to collect the necessary data from a participant if traditional procedures for contingency assessment were used. This has led Allan et coll. (Crump, Hannah, Allan, & Hord, 2007; Hannah, Crump, Allan, & Siegel, 2009; Siegel et al., 2009) to design the streamed-trial procedure. In that procedure, participants are exposed to very rapid streams of stimuli (each stimulus lasts about 100 ms) before having to judge the contingency between the target cue and the target outcome. The short duration of the stimuli allows researchers to collect all the data necessary for the SDT analysis while easily keeping the duration of a session well under an hour.

Using this procedure, Maia et al. (2018) studied how the sensitivity to the contingencies (mean M of the distributions of outcome activation) and the variability in the outcome activation (standard deviation D of the distributions of outcome activation) varied as a function of ΔP. They found that participants were more sensitive to the positive contingencies than to the negative ones while the variability in the outcome activation remained roughly constant across ΔP values even though it tended to increase slightly as ΔP became more extreme. Another noteworthy result was the huge range of inter-individual differences observed by Maia et al. (2018): some participants were extremely sensitive to variations in ΔP while others almost displayed a form of contingency blindness. But, as the participants only came once to the laboratory, it is not possible to know whether those individual differences were stable across time: if a participant who was extremely sensitive to variations in the contingencies had come back for a second experimental session, would she have still displayed a strong sensitivity to the contingency? Likewise, for a participant who displayed quasi-contingency blindness? Moreover, in the Maia et al. (2018)'s study, the participants were not given any feedback regarding their judgment. Would their sensitivity to the contingencies improve if they had been provided with such a feedback? Or would mere repeated exposure to the task suffice to improve their performance?

The goal of this study was to provide answers to these simple questions. Participants were exposed repeatedly to the streamed-trial procedure. The parameters of the SDT model were recovered from these repeated sessions in order to assess whether they remained stable across sessions or not. Moreover, during some sessions, the participants received feedback about their performance in order to see whether this would impact the SDT parameters.

Method

Participants and apparatus

4 participants (1 male, 3 females, age ranging from 20 to 26 years old), all students at the University of Lille, were recruited for this study. They were all major in either psychology or cognitive science. The program for the study was written in Python using the Psychopy2 library (Peirce, 2007). The participants ran this program at home on their own computer. They were instructed to try to do one session at least 5 days a week at a moment where they would feel rested. They were told to try to sit about 60 cm from the screen.

Procedure

During a session, a participant was exposed to rapid streams of stimuli. A stream started with a 1-s black screen with a fixation cross at its center. The cross remained visible until the end of the stream. The fixation cross was followed by 20 trials. Each trial lasted for 200 ms and comprised two phases: (a) the cue phase lasting 100 ms in which a yellow triangle (stimulus A) was always displayed in the upper right-hand corner of the screen, marking the beginning of a trial. On some trials, a green circle (target cue X) was displayed in the upper left-hand corner of the screen; (b) the outcome phase lasting 100 ms. In some trials, a red square (outcome O) was presented in the lower central part of the screen. All stimuli measured 500 by 500 pixels. Trials were separated by a 100-ms inter-trial interval (ITI) during which only the fixation cross was displayed.

At the end of a stream, a dialog box appeared at the center of the screen displaying the sentence: "The contingency between the circle and the square was…" Three boxes appeared below reading “Negative” (on the left), “Null (centered), and “Positive” (on the right). The participants understood that they should answer “Negative” if they thought the square appeared less often when the circle was presented than if it was not presented, “Null” if they thought the square appeared as often whether the circle was presented or not, and “Positive” if they thought the square appeared more often when the circle was presented than if it was not presented. The participants used the mouse to answer (the pointer always appeared centrally, as equidistant as possible from the three response options). Once they clicked on one of the buttons, the dialog box disappeared and another one appeared with the question “How sure are you of your decision?”. If they had clicked on either “Negative” or “Positive”, the two buttons appearing below the confidence rating question and on which they could answer by clicking on them read, “Not sure” (appearing centered right), and “Sure” (centered). If they had clicked on “Null”, three buttons below the confidence rating read (in that order, from left to right), “Not sure, could be negative”, “Sure”, and “Not sure, could be positive”. In the No-Feedback condition, once the participant answered this question, the dialog box disappeared and a new stream started, except if this was the end of a session, in which case the program ended. Things proceeded in a similar way in the Feedback condition, except that, if the participant had correctly identified the contingency, a screen reading “Correct!” appeared for 1 second before the start of the next stream (or the program ending if this had been the last stream). Participants were aware that a positive feedback was systematically given when their response was correct, and hence, that an absence of feedback meant that their response was incorrect.

Overall, a stream was composed of four types of trials: (a) AX+ trials in which both A and X were shown during the cue phase, the outcome being shown during the outcome phase; (b) AX- trials in which both A and X were shown during the cue phase, the outcome not being shown during the outcome phase; (c) A+ trials in which only A was shown during the cue phase, the outcome being shown during the outcome phase; (d) A- trials in which only A was shown during the cue phase, the outcome not being shown during the outcome phase.

Manipulating the proportion of these four types of trials within a stream induced different ΔP values between the target cue X and the outcome. During a session, the participant was exposed to 3 types of streams: negative streams (ΔP = -0.4: 3 AX+, 7 AX-, 7 A+, 3 A-), null streams (ΔP = 0: 5 AX+, 5 AX-, 5 A+, 5 A-), and positive streams (ΔP = 0.4: 7 AX+, 3 AX-, 3 A+, 7 A-). Note that the probability of the cues and outcomes was kept constant across streams as well as the number of cue and outcome presentations. Each stream type was presented 60 times during a session. The order of presentation of the streams was determined randomly (without replacement) just like the order of presentation of the trials every time a stream was presented.

The procedure used an ABA single-subject design (Sidman, 1960). The participants were first exposed to the No-Feedback condition for a minimum of 15 sessions. The data were emailed to the experimenter as soon as they were collected. The experimenter inspected them visually to decide whether performance was stable enough to switch the participant to the Feedback condition. The same way, once that transition took place, the participants were exposed to the Feedback condition for a minimum of 15 sessions. If the visual inspection of the data revealed that the performance was stable, the participant was switched back to the No-Feedback condition in which sessions continued until the participant was not available anymore to participate to the study (usually because the end of the term had been reached).

Data analysis

Data analysis proceeded according to the template provided by Maia et al. (2018) and Jozefowiez, Gaudichon, Mekkass, & Machado (2018). Figure 1 shows how the experimental task can be conceptualized using SDT. There are three distributions of outcome activation, each corresponding to a ΔP condition and 6 criteria delimitating 7 areas corresponding to the various responses the participant can provide during a session ("Negative - Not Sure", "Negative - Sure", "Null - Not Sure, could be negative", "Null - Sure", "Null - Not Sure, could be positive", ""Positive - Sure", "Positive - Not Sure").

Let Z be the inverse of the cumulative standard normal distribution. For each participant and for each ΔP condition i (i = [-0.4, 0, 0.4]), we can compute $Z(p_i) = \{Z[p_i(C1)], Z[p_i(C2)], Z[p_i(C3)], Z[p_i(C4)], Z[p_i(C5)], Z[p_i(C6)]\}$: $p_i(C1)$ is the probability for the participant to respond "negative – not sure", "null" (at any level of confidence) and "positive" (at any level of confidence) after being presented with a stream in which ΔP was equal to i; $p_i(C2)$ is the probability for the participant to respond "null" (at any level of confidence) and "positive" (at any level of confidence) after being presented with a stream in which ΔP was equal to i; $p_i(C3)$ is the probability for the participant to respond "null –sure/not sure, it could be positive" after being presented with a stream in which ΔP was equal i; etc.

By plotting $Z(p_0)$ vs. $Z(p_i)$ (i = [-0.4,0.4]), we obtain two zROC curves. Let's assume that when the participant is presented with a stream in which ΔP = 0 (respectively ΔP = i), the outcome activation is drawn from a Gaussian distribution with mean μ_0 (respectively μ_i) and standard deviation σ_0 (respectively σ_i). It can be shown (see Wickens, 2002 or Appendix I of Jozefowiez et al, 2018 for details) that

$$Z(p_i) = \frac{\sigma_0}{\sigma_i} Z(p_0) + \frac{\mu_i - \mu_0}{\sigma_i} \qquad (1)$$

By setting μ_0 and σ_0 to arbitrary values, it is possible in theory to retrieve the values of μ_i and σ_i (relative to μ_0 and σ_0) by fitting Equation (1) to the zROC curve.

In practice, fitting a straight line to a z-ROC curve is not so straightforward. Standard fitting techniques assume that the predicting variable is an independent variable (so its value is known without any error) while the predicted variable is a dependent variable (for which error in measurement are possible). In zROC curves, both the predicting and the predicted variable are dependent variables. Hence, standard linear regression cannot be used and iterative maximum-likelihood algorithms must be used instead (Macmillian & Creelman, 1991; Wickens, 2002). To perform those computations, I used the RScorePlus program, designed by Dr. Lewis O. Harvey Jr. from the University of Colorado and available at http://psych.colorado.edu/~lharvey/html/software.html. The way I used it, the program performed the following steps: (a) It added 1/7 to each rating data. This is the so-called loglinear correction where 1/m (m being the number of possible responses given by the participant) is added to all the data to avoid ratings with a frequency of 0, that would be problematic in the analysis (notably because $Z(0)$ is not defined. See Hautus & Lee, 1998); (b) Setting μ_0 to 0 and σ_0 to 1, the program fitted Equation (1) to the z-ROC curves using least-square linear regression techniques in order to get good starting values for the 10 parameters of the SDT model (the mean and standard deviations of the 2 Gaussian distributions corresponding to the 2 remaining ΔP conditions plus the values of the 5 decision criteria $C1, C2, \ldots, C6$. $Z[p_0(C1)] = C1, Z[p_0(C2)] = C2$, etc. See Wickens, 2002); (c) Using those estimates as the input of the Marquardt nonlinear least-square procedure (Marquardt, 1963; Press, Teukolsky, Vetterling, & Flannery, 2007), the program found the maximum-likelihood estimates of the 10 parameters.

Note that the null contingency was chosen as the reference condition purely out of convenience: by setting its mean to 0, the positive contingency distribution would have a positive mean while the negative contingency would have a negative mean. This makes the psychological interpretation of the parameters easier. If the positive contingency distribution or the negative contingency one had been chosen instead as reference, this would have not changed the conclusions: it would have only impacted the absolute values of the SDT parameters, not their relative ones.

Based on the likelihood computed by the program that the data could have been generated by the SDT model, a chi square was computed to quantify the fit between the model and the data. A significant chi square at the conventional threshold of α = .05 was interpreted as indicating a potential discrepancy between the model and the data.

When necessary, 95% confidence interval (CI) are reported. They are computed using Student's *t* distribution. Likewise, when necessary, Cohen's *d* (more precisely Hegde's *g*, which is a bias-free estimate of the population Cohen's *d*) is provided to measure effect size between two conditions 1 and 2 (Cummings, 2012). 95% CI for Cohen's *d* were computed using ESCI (https://thenewstatistics.com/itns/esci/) which implements the method described in Cummings (2012).

<u>Results</u>

Table I shows the number of sessions a participant spent in a condition. The top of Figures 2 to 5 show for each participant and across sessions the proportion of streams for which the participant identified a specific contingency (negative, null, or positive) as a function of the actual cue-outcome contingency and of the condition (no feedback vs. feedback). The bottom of Figures 2 to 5 shows for each participant and across sessions the proportion of streams in which the participant identified a specific contingency (negative, null, or positive) while being sure of her answer, as a function of the objective cue-outcome contingency (ΔP = -0.4, 0, or 0.4) and the condition (no feedback vs. feedback). These data are provided for reference only. The goal of the SDT analysis is to disentangle from them the contribution of the multiple factors (sensitivity to the contingencies, that is to say the mean of the distributions of outcome activation; variability in outcome activation, that is to say the standard deviation of the distributions of outcome activation; decision criteria) contributing to the performance of the participants.

Fit of the SDT model.

Figure 6 shows the *p*-values for the chi square measuring the fit of the SDT model to the data across sessions and conditions for each of the participants. Nothing in this figure suggests that the SDT model had difficulty accounting for the participants' performance in any of the conditions, with the possible exception of participant P3. For instance, in the second no-feedback condition, there was a significant difference between the model and the data in 10 out of 22 sessions. The reasons why there was such a divergence are unclear. Figure 7 shows the average ratings provided by P3 as a function of the objective contingency for the 12 sessions in the second no-feedback condition for which there was no significant difference between the SDT model and the data (top panel) and the 10 sessions for which there was such a significant difference (bottom panel). The average predictions of the SDT model are also reported on those graphs. As can be seen, they do not seem better or worse when a significant difference between the data and the model was detected than when such a difference was not detected. Hence, it seems that the significant chi-square was not due to some systematic deviation between the prediction of the SDT model and the performance of the participant. Maybe the participant answered more randomly because of a momentary lack of attention during the sessions for which a significant difference was detected: this would have caused some difficulty for the SDT model to perfectly account for her performance. Still, this did not impact the estimate of the parameters of the SDT model: as can be seen in the graphs that will be shown subsequently in this article, the sessions for which a significant difference between the model and the data was detected do not stand out from the sessions for which there was no such difference as far as the parameters of the SDT model are concerned.

Sensitivity to the contingencies: mean of the distributions of outcome activation

Figure 8 shows the mean of the distributions of outcome activation. Despite the small sample size, individual differences in the sensitivity to the contingencies can still be observed in this study. If we use the average sensitivity to the contingencies during the first no-feedback condition as a criterion, the performance of P2 and P3 are very similar. They both display a very good sensitivity to the contingencies (average sensitivity to the negative contingency: P2, -1.23, 95% CI [-1.36, -1.11]; P3, -1.17, 95% CI [-1.24, -1.11]; average sensitivity to the positive contingency: P2, 1.14, 95% CI [0.98, 1.29]; P3, 1.13, 95% CI [1.01, 1.25]). P1's performance is on the other hand quite poor, more so for the negative contingency than for the positive one (average sensitivity to the negative contingency: P1, -0.60, 95% CI [-0.73, -0.46]; average sensitivity to the positive contingency: P1, 0.69, 95% CI [0.56, 0.83]). Indeed, as shown in Figure 2, this participant almost always identified a null contingency regardless of the actual cue-outcome contingency. Participant P4 falls between P1 on one hand, and P2 and P3 on the other: she is on par with P1 regarding her sensitivity to the negative contingency (average sensitivity to the negative contingency: P4, -0.64, 95% CI [-0.72, -0.56]) but she is better at perceiving the positive contingency even though her performance remains below the one of P2 and P3 (average sensitivity to the positive contingency: P2, 0.90, 95% CI [0.83, 0.96]).

With the exception of P1 once feedback is introduced, there is very little variation in the sensitivity to the contingencies inside a condition. Hence, the average sensitivity to the contingencies during a condition provides an accurate summary of the data. They are displayed in Figure 9. Combining this figure with Figure 8 allows us to reach the following conclusions.

First, the introduction of the feedback had a spectacular effect on the sensitivity to the contingencies for P1. Her sensitivity to both the positive and the negative contingency rapidly increased once she received feedback about her performance until reaching a plateau. This plateau was reached faster for the positive contingency than for the negative one (first non-feedback condition vs. feedback condition: positive contingency, Cohen's $d = 2.91$, 95% CI [2.06, 3.86]; negative contingency, Cohen's $d = -2.40$, 95% CI [-3.26, -5.60]. Note that because P1's performance does not reach its asymptote in the feedback condition immediately, these effect sizes actually underestimate the impact of feedback on P1's performance. A more accurate estimate can be obtained by comparing the first non-feedback condition to the second non-feedback condition: positive contingency, Cohen's d = 3.83, 95% CI [2.77, 5.04]; negative contingency, Cohen's d = -4.29, 95% CI [-5.60, -3.14]). As a consequence, her sensitivity to the contingencies by the end of the study is higher than any of the other participants that outmatched her at the beginning of the study. Moreover, this improvement in her sensitivity to the contingencies is not affected by the discontinuation of the feedback.

The performance of P3 is also affected by the feedback but in a less dramatic way. There is an increase in her sensitivity to the positive contingency (first non-feedback condition vs. feedback condition: positive contingency, Cohen's d = 0.79, 95% CI [0.29, 1.30]) and a much larger increase in her sensitivity to the negative contingency (Cohen's d = -1.10, 95% CI [-1.63, -0.58]). In both cases, the change in sensitivity is not progressive as in the case of P1: performance reached its asymptote in the first session in which the feedback is introduced. P3's increased sensitivity to the positive contingency resisted the discontinuation of feedback (feedback condition vs. second non-feedback condition: Cohen's d = 0.31, 95% CI [-0.24, 0.86]; first non-feedback condition vs. second non-feedback condition: Cohen's d = 1.29, 95% CI [0.71, 1.89]). This is not the case of the increase in her sensitivity to the negative contingency: it went back to its baseline level once feedback was discontinued (feedback condition vs. second non-feedback condition: Cohen's d = 0.72, 95% CI [0.17, 1.3]; first feedback condition vs. second non-feedback condition: Cohen's d = -0.34, 95% CI [-0.88, 1.19]).

On the other hand, the feedback does not seem to have had an impact on the performance of either P2 nor P4. Yet their performance did change across the experiment. P2's performance seems to have gotten worse: her sensitivity to the contingency progressively decreased during the study, more so for the negative contingency than for the positive one (first non-feedback condition vs. second non-feedback condition: positive contingency, Cohen's *d* = -0.40, 95% CI [-0.88, 0.08]; negative contingency, Cohen's *d* = 0.81, 95% CI [0.31, 1.31]). Quite on the contrary, P4's performance improved: her sensitivity to both the positive and the negative contingency increased during the experiment (first non-feedback condition vs. second non-feedback condition: positive contingency, Cohen's *d* = 1.01, 95% CI [0.44, 1.57]; negative contingency, Cohen's *d* = -1.11, 95% CI [-1.78, -0.55]).

Variability in outcome activation: standard deviation of the distributions of outcome activation

Figure 11 shows the standard deviation of the distributions of outcome activation for each participant across sessions and as a function of conditions. There is no systematic change in this measure during a condition. Hence, averaging it over a condition provides an accurate summary of the data. The average standard deviation across conditions for each participant is shown in Figure 11. Combining this figure with Figure 10, the following conclusions can be reached.

First, there is no detectable difference in the standard deviation between conditions for any of the participants. For P1 and P4, the variability in outcome activation is larger for the positive and the negative contingencies than for the null contingency whose standard deviation in the SDT model is set by convention to 1. This is also the case for P2 and P3 as far as the negative contingency is concerned. Moreover, for these two participants, outcome activation is much more variable for the negative contingency than for the positive one while there is no such difference for P1 and P4.

Decision criteria

Figure 12 shows the decision criteria for each participant across sessions and conditions. Those criteria are remarkably stable though they do sometimes evolve. Notice, for instance, that P4's criteria shift downward at the very beginning of the experiment. P1's change in criteria are clearly linked to the introduction of feedback and to its discontinuation.

Initially, the criteria are set so that a large range of outcome activation values led P1 to decide that the cue-outcome contingency was null (see also Figure 1). The criteria are revised once feedback is introduced: the perception of a null contingency is now tied to a much narrower set of outcome activation values. Note also that, contrary to the change in the sensitivity to the contingency which was progressive for that participant (Figure 8), the change in criteria is immediate, taking place after the introduction of the feedback with no further change until the end of the condition. Once feedback is discontinued, the criteria do not reverse back to their older values but a further change occurs: mainly, the participant is less sure of her answer when she categorizes a stream as either implementing a positive or a negative contingency.

The changes in criteria for the other participants are more difficult to link to the introduction or discontinuation of the feedback. P4's changes in criteria take place gradually across the first condition before stabilizing: it is clearly unrelated to the presence of feedback. The range of values leading P2 to identify a negative contingency increases at the end of the feedback condition but similar variations are observed when feedback is discontinued.
As for P3, the feedback seems to make her slightly more confident that she identified a positive or a negative contingency while the discontinuation of the feedback is shifting upward all of her criteria.

General Discussion

The questions asked by this study were modest in scope but, I believe, worth asking: is the ability of a participant to discriminate between cue-outcome contingencies stable over time? Can this ability be improved if feedback is provided to the subject? To answer these questions, the present experiment provided a detailed case study of the performance of 4 participants. The use of SDT analysis allowed to examine separately the various components contributing to performance (sensitivity to the contingencies, variability in outcome activation, decision criteria).

Synthesis

The ability of the participants to discriminate between the contingencies relies on a participant's sensitivity to the contingencies (mean of the distributions of outcome activation) and the variability in outcome activation (standard deviation of the distributions of outcome activation). Concerning the sensitivity to the contingencies, spontaneous changes were observed in two participants (one improved over time, the other deteriorated) but these changes were progressive and only detectable over the long term. The variability in outcome activation remained on the other hand constant for all the participants, with no detectable increasing or decreasing trends. Finally, decision criteria also changed during the study, though, overall, it is much more remarkable how stable they were.

Of course, one could argue with the definition of "stable over time" used in the present study. The participants were studied over a period of month which, as far as experiments in the field of learning with human participants are concerned, is quite a long period. But it is, of course, ridiculously short compared to most longitudinal studies which can follow participants over a period of years. If the participants had come back a year later to the lab, would their performance be similar to the one observed at the end of this study? Notably, would the spectacular improvement in performance caused by feedback in participant P1 still be there or would she had reverted to her performance level before the introduction of feedback? All those questions are interesting but, unfortunately, it is impossible to provide answers to them for obvious practical reasons.

Concerning the second question, the feedback is definitely able to alter both a participant's sensitivity to the contingencies and the decision criteria used by that participant though the effect differs depending on the participant. P1's sensitivity to the contingencies and decision criteria were strongly impacted by the feedback. The effect was more subdued on P3's sensitivity to the contingencies. On the other hand, this participant's decision criteria were not affected by the feedback. There was no effect of the feedback on P2 and P4's sensitivity to the contingencies and it is dubious whether the change in criteria observed in P2 can be traced to the introduction of the feedback. Likewise, participants who showed an effect of the feedback differed regarding how they responded to the discontinuation of the feedback: P1's sensitivity to the contingencies was not impacted while only P3's sensitivity to the positive contingency did not reverse to its baseline level; both participants' criteria were affected by the discontinuation of the feedback even though its introduction had little impact on P3's performance. On the other hand, the variability in outcome activation was not impacted neither by the introduction of the feedback nor by its discontinuation.

As P2 and P4 were the two participants whose sensitivity to the contingencies was the highest at the beginning of the study (though P1 outmatched them by the end of the study), it would be tempting to attribute the lack of effect of feedback on their sensitivity to the contingencies to a ceiling effect. But, as Figures 3 and 5 shows, there is still room for improvement for those two participants: thought they are quite good at identifying the positive streams, they do not identify correctly a lot of the null and negative ones. Still, the huge amount of negative feedback they must have received after a null or a positive stream as a consequence did not lend to a change in performance.

On a side note, the data nicely illustrates the discrepancy between the participant's performance (captured notably by their sensitivity to the contingencies) and their subjective perception of the contingencies (captured by the decision criteria). Take P1 as a case study. During the first non-feedback condition, she usually perceived all cue-outcome contingencies as null with a high degree of confidence, even when the actual cue-outcome contingency was positive or negative. She actually reported that she was unable to do the task as she could not see the differences between the various streams. Despite this, the SDT analysis revealed that she was indeed sensitive to the contingencies. Once feedback was introduced, her sensitivity to the contingencies improved greatly and she also became more confident when perceiving positive and negative contingencies. It is noteworthy that this effect on the decision criteria was immediate while the improvement in her sensitivity to the contingencies was only gradual. Likewise, when feedback was discontinued, she became less confident when perceiving positive and negative contingencies despite the fact that her sensitivity to the contingencies remained the same compared to the feedback condition (as a side note, P3 was unable to explain her improvement in performance once the feedback was introduced. She just remarked that, while the task seemed difficult before, it now seemed easier though she did not know why).

Individual versus group performance

These individual differences stand out because the study used a single-subject design. The use of such methodology is standard in the field of operant conditioning which emphasizes individual performance (Sidman, 1960). It is also often used in SDT studies: instead of studying a large number of participants for a small number of sessions, a small number of expert participants is studied for a large number of sessions (Macmillian & Creelman, 1991; Wickens, 2002). Single-subject design are much more unusual in the field of Pavlovian conditioning and contingency assessment which, as most of psychology, relies on statistical analysis of group averages. Both methods have their advantages and their inconveniences. One problem with the group average method is that it is unable to detect participants for which the effects detected at the group level do not apply or vice versa. For instance, if based on a group average, the conclusion of the present study would have been that the feedback is able to change the sensitivity of the participants, foregoing the fact that it had no impact on the sensitivity to the contingencies for P2 and P4. Or, if more participants had been run and most of them were like P2 and P4, the conclusion would have been that the sensitivity to the contingencies cannot be altered by feedback, bypassing the fact that feedback had a massive effect for some participants such as P1. Because conclusions on group average cannot detect this kind of difference between participants, researchers tend to overlook them. This leads to a kind of data blindness which limits our theories as we only try to account for the effect detected at the group level, overlooking individual differences and failing to account for them (see Jozefowiez et al., 2018 for a similar argument with regard to interval timing).

One virtue of the present study is that it documents individual differences in the way participants process cue-outcome contingencies. One of its limits is that it cannot explain them. This would require further research foregoing the group analysis techniques favored by research on human associative learning to focus more on single-subject performance analysis. Notably, it would be interesting to know why the feedback had such a tremendous effect on some participants while it had none on others.

Further limits of the current study can be easily identified. Only a small number of participants can participate to this kind of single-subject design, as it requires finding participants who will willingly accept to perform the experimental task over a long period of time. In this case, the participants were motivated by the fact that participating to the project allowed them to partially fulfill a class requirement. Other times, participants might receive a financial compensation. In both cases, this means that resources available for recruiting participants are limited and do not allow for a large sample size. This precludes investigations of the impact of variables such as gender, age, or education level. The 4 participants who took part of the current study all had approximatively the same age and the same education level. As for gender, there is no obvious difference between the three female participants and the single male one (P2). A better strategy to know whether those variables might have an impact on performance might be to better understand the cognitive processes at work. If we independently know whether those processes are affected by gender, age, or similar variables, we would know likewise that these variables would affect contingency assessment.

Another limit of the present study is that the range of variables whose impact on performance was studied was rather limited. Besides feedback, we could have looked at the impact of attention and motivation, for instance. The former is the topic of a yet unpublished experiment. This was why I decided not to manipulate it in the current study. Motivation would have been hard to manipulate as financial incentives could not have been used as the participants were students taking part to the project in partial fulfillment of a course requirement.

Comparison with Maia et al. (2018)

This study was a follow-up to Maia et al. (2018) which used more of a group approach. Hence, it would be interesting to see whether the conclusion of the present study are in line with Maia et al. (2018) despite the use of a single-subject design. Maia et al. (2018) highlighted two conclusions in their study: the participants were more sensitive to the positive contingency than to the negative ones; the higher the absolute value of ΔP, the more variable the outcome activation (which they call the "subjective contingency"). Both conclusions are reflected in the present study but in a different way.

If we look at the average sensitivity to the contingencies during a condition, it is not obvious that the participants are more sensitive to the positive contingency: P1 and P4 are; P2 is actually more sensitive to the negative contingency; P3 seems to be overall as sensitive to both. This is not inconsistent with Maia et al. (2018)'s conclusions as theirs were based on group average. Even with only 4 participants, the same conclusion would have been reached here if the data from all the participants had been averaged. Yet, there are other indications within the data that the participants had more difficulty processing negative contingencies than positive ones: once feedback was introduced, P1 reached its asymptotic sensitivity faster for the positive contingency than for the negative contingency; the feedback had a greater impact on P3's sensitivity to the positive contingency than on her sensitivity to the negative contingency; once feedback was discontinued, P3's sensitivity to the negative contingency returned to her baseline level while this was not the case for her sensitivity to the positive contingency; for P2 and P4, the variability in outcome activation was higher for the negative contingency than for the positive one. Overall, these arguments pile up to the conclusion that processing positive contingencies is slightly easier for human participants than processing negative ones.

Concerning the variability in outcome activation, the present data are clearly in line with the one reported by Maia et al. (2018). Except possibly for P2 and P4 when the positive contingency is concerned, outcome activation is clearly more variable if the contingency is positive or negative than if it is null. Once again, if the data had been averaged, this would have been the conclusion that would have been reached.

Conclusion

Overall, the present study further illustrates the value of a signal detection approach to associative learning. It reveals that the sensitivity to the contingency is a stable feature of a participant: some are very good at it, other ones less so. It also indicates that, while feedback can improve a participant's performance, individuals differ regarding their susceptibility to such an intervention. This conclusion calls for a renewed focus on individual performance in the study of associative learning.

References

Allan, L. G. (1993). Human contingency judgments: Rule-based or associative? *Psychological Bulletin, 114,* 435-448.

Allan, L. G., Hannah, S. D., Crump, M. J. C., & Siegel, S. (2008). The psychophysics of contingency assessment. *Journal of Experimental Psychology: General, 137,* 226-243.

Allan, L. G., Siegel, S., & Tangen, J. M. (2005). A signal detection analysis of contingency data. *Learning & Behavior, 33,* 250-263.

Crump, M. J. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgments on the fly. *Quarterly Journal of Experimental Psychology, 60,* 753-761.

Cummings, G. (2012). *Understanding the new statistics: effect size, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Hannah, S. D., Crump, M. J. C., Allan, L. G., & Siegel, S. (2009). Cue-interaction effects in contingency judgments using the streamed-trial procedure. *Canadian Journal of Experimental Psychology, 63,* 103-112.

Harvey, L. O. (2013). *Parameter estimation of signal detection models: RscorePlus user's manual* (Tech. Rep.). Boulder, Colorado 80309-0345 USA http://psych.colorado.edu/~lharvey/Software%20Zip%20Files/RscorePlus_doc_5_8_3.pdf. Department of Psychology and Neuroscience University of Colorado.

Hautus, M. J., & Lee, A. J. (1998). The dispersions of estimates of sensitivity obtained from four psychophysical procedures: implications for experimental design. *Perception & Psychophysics, 60*, 638-649.

Jozefowiez, J., Gaudichon, C., Mekkass, F., & Machado, A. (2018). Log versus linear timing in human temporal bisection: A signal detection study. *Journal of Experimental Psychology: Animal Learning & Cognition, 44,* 396-408.

Laux, J. P., Goedert, K. M., & Markman, A. B. (2010). Causal discounting in the presence of a stronger cue is due to bias. *Psychonomic Bulletin & Review, 17,* 213-218.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276-278.

Macmillian, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics, 11,* 431-441.

Maia, S., Lefèvre, F., & Jozefowiez, J. (2018). Psychophysics of associative learning: Quantitative properties of subjective contingency. *Journal of Experimental Psychology: Animal Learning & Cognition, 44,* 67-81.

Miller, R. R. & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed), *The Psychology of Learning and Motivation* (Vol. 22, pp. 51-92). San Diego, CA: Academic Press.

Pearce, J. M. (1987). A model of generalization for Pavlovian conditioning. *Psychological Review, 94,* 61-73.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review, 87,* 532-552.

Perales, J. C., Catena, A., Shanks, D. R., & González, J. A. (2005). Dissociation between judgments and outcome-expectancy measures in covariation learning: A signal detection theory approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1105-1120.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical recipes in C++: The art of scientific computing (2nd edition).* Cambridge: Cambridge University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes: The art of scientific computing (3rd edition).* New York: Cambridge University Press.

Peirce, J. W. (2007). Psychopy: Psychophysics software in Python. *Journal of Neuroscience Methods, 162,* 8-13.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & F. Prokasy (Eds), *Classical Conditioning II* (pp. 64-99). New York: Appleton Century Crofts.

Sidman, M. A. (1960). *Tactics of Scientific Research.* New York: MacMillan.

Schanks, D. R. (1995). *The psychology of associative learning.* Cambridge: Cambridge University Press.

Schanks, D. R. (2007). Associationism and cognition: Human contingency learning at 25. *Quarterly Journal of Experimental Psychology, 60,* 291-309.

Siegel, S., Allan, L. G., & Hannah, S. D. (2009). Applying signal detection theory to contingency assessment. *Comparative Cognition & Behavior Reviews, 4,* 116-134.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds), *Information processing in animals: Memory mechanisms* (pp. 5.47). Hillsdale, NJ: Erlbaum.

Wickens, T. D. (2002). *Elementary signal detection theory.* New York: Oxford University Press.

Author Notes

Jérémie Jozefowiez, Univ.Lille, CNRS, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, F-59000 Lille, France. Correspondence concerning this article should be addressed to Jérémie Jozefowiez, Laboratory of Cognitive and Affective Sciences (SCALab UMR CNRS 9193), Université de Lille, Domaine Universitaire du Pont de Bois, BP 60149, 58653 Villeneuve d'Ascq Cedex, France. E-mail: jeremie.jozefowiez@univ-lille.fr. All raw data and computer code for the experiment is available upon request from J. Jozefowiez. The author reports no conflicts of interest.

Table I. Number of sessions in a condition for each participant

Participant	No Feedback (A)	Feedback (B)	No Feedback (C)
P1	21	20	15
P2	35	49	32
P3	34	31	22
P4	44	38	19

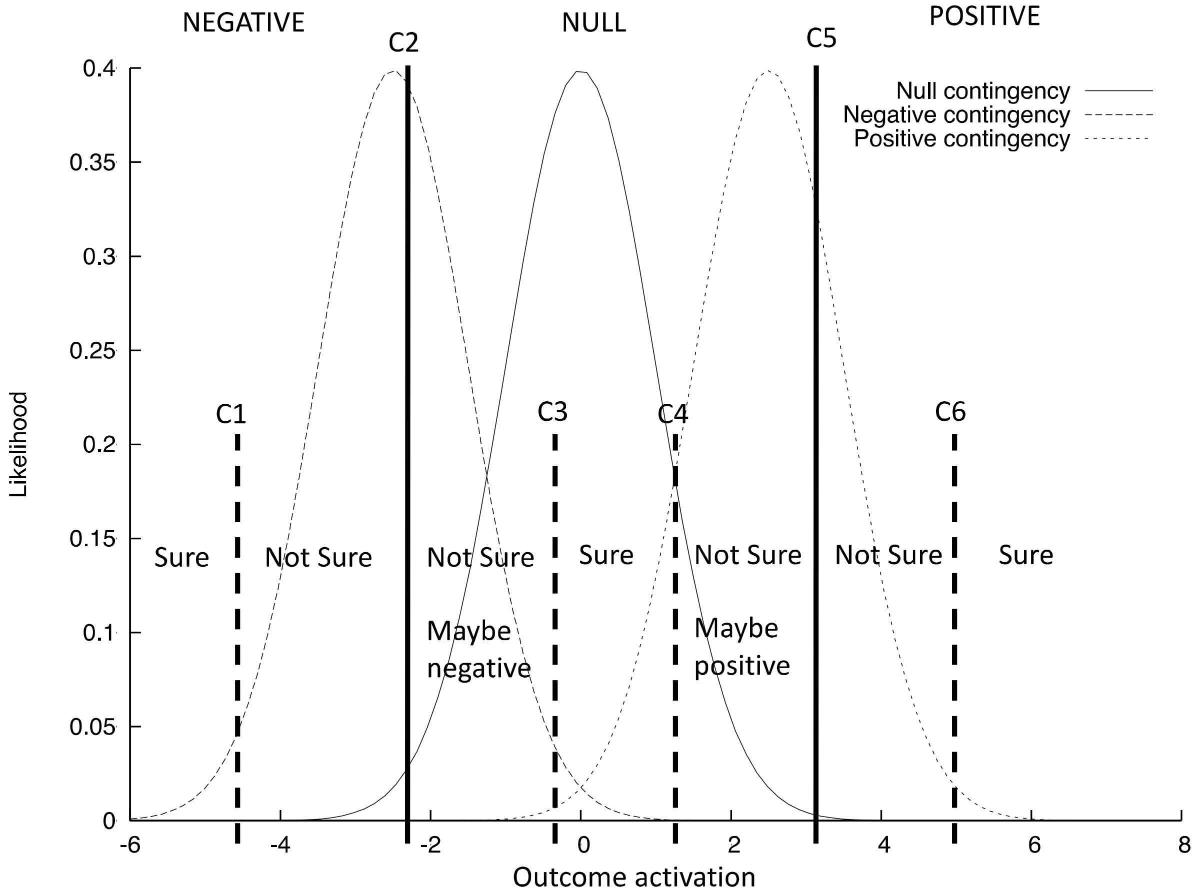
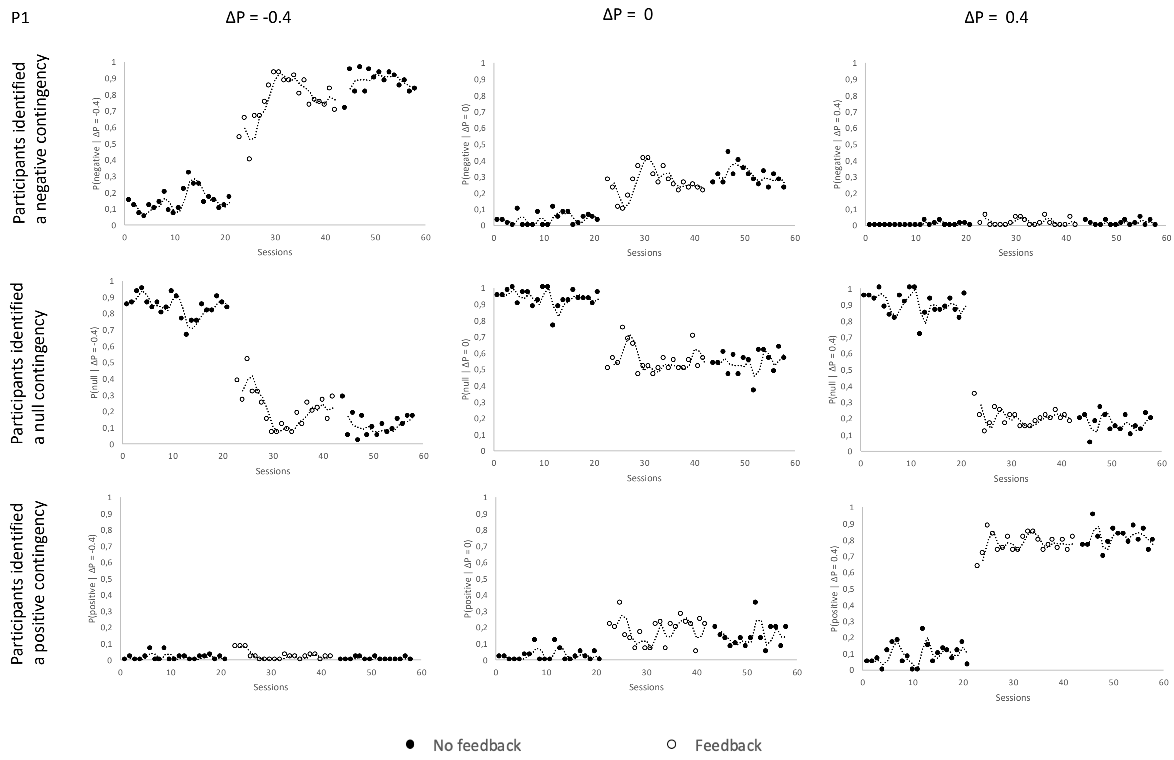


Figure 1. Signal detection model of the streaming task. At the end of a stream, outcome activation is a random variable drawn from a Gaussian distribution whose mean and standard deviation are a function of the objective contingency between the cue and the outcome. The participant perceives a negative contingency if the outcome activation is below C2; she is sure of that if the outcome activation is below C1 but not sure if it is between C1 and C2. The participant perceives a null contingency if the outcome activation is between C2 and C5: she is sure of that if it is between C3 and C4 but unsure if it is either between C2 and C3 (the contingency might be negative) or between C4 and C5 (the contingency might be positive). Finally, the participant perceives a positive contingency if the outcome activation is larger than C5: she is unsure of that if it is between C5 and C6, and sure if it is greater than C6.



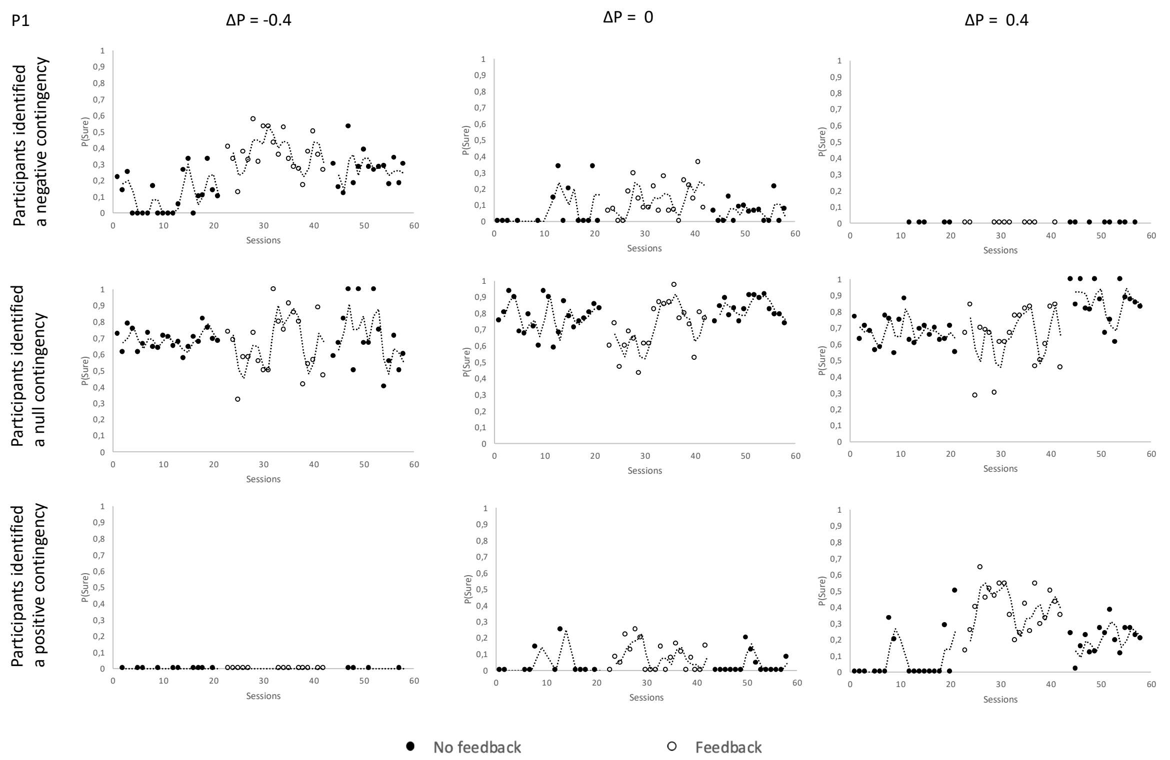


Figure 2. Top: Probability to identify a specific contingency (positive, null, or negative) as a function of the objective contingency across sessions and conditions for participant P1. Bottom: Proportion of trials for which participant P1 identified a specific contingency

(positive, null, or negative) while claiming to be sure as a function of the objective contingency across sessions and conditions. The dashed line is an exponential moving average with a window of 2.

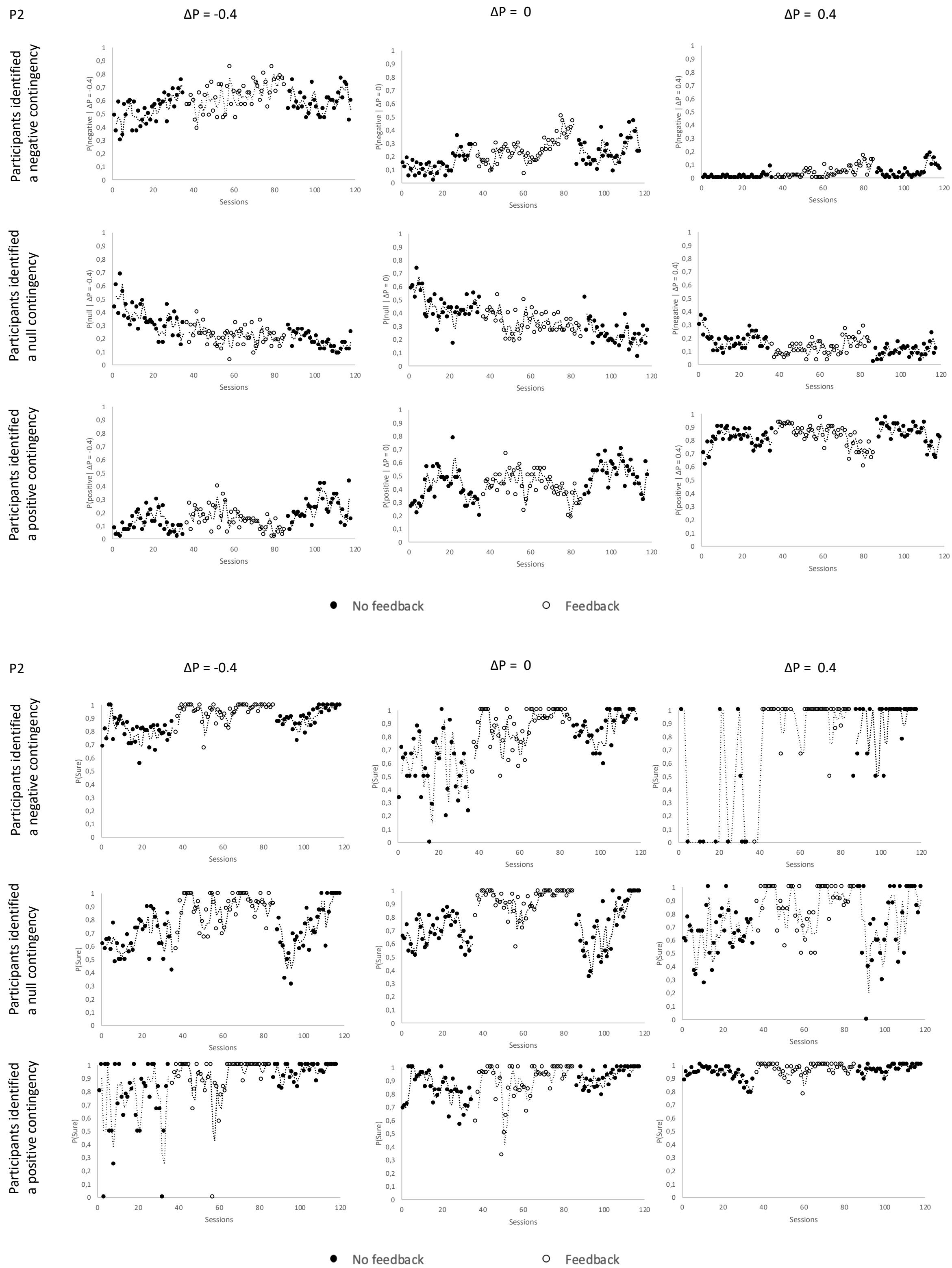


Figure 3. Top: Probability to identify a specific contingency (positive, null, or negative) as a function of the objective contingency across sessions and conditions for participant P2. Bottom: Proportion of trials for which participant P2 identified a specific contingency

(positive, null, or negative) while claiming to be sure as a function of the objective contingency across sessions and conditions. The dashed line is an exponential moving average with a window of 2.

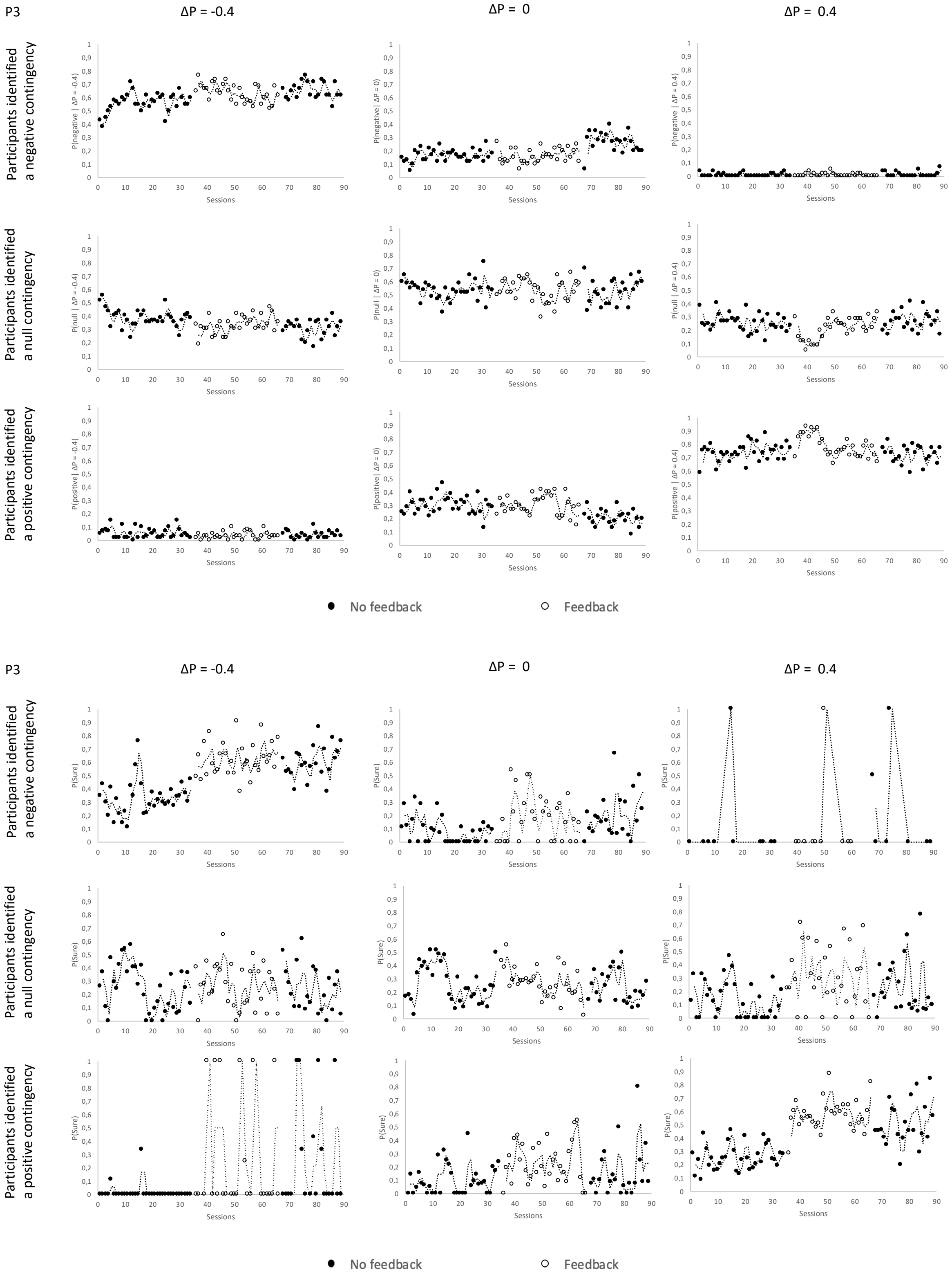
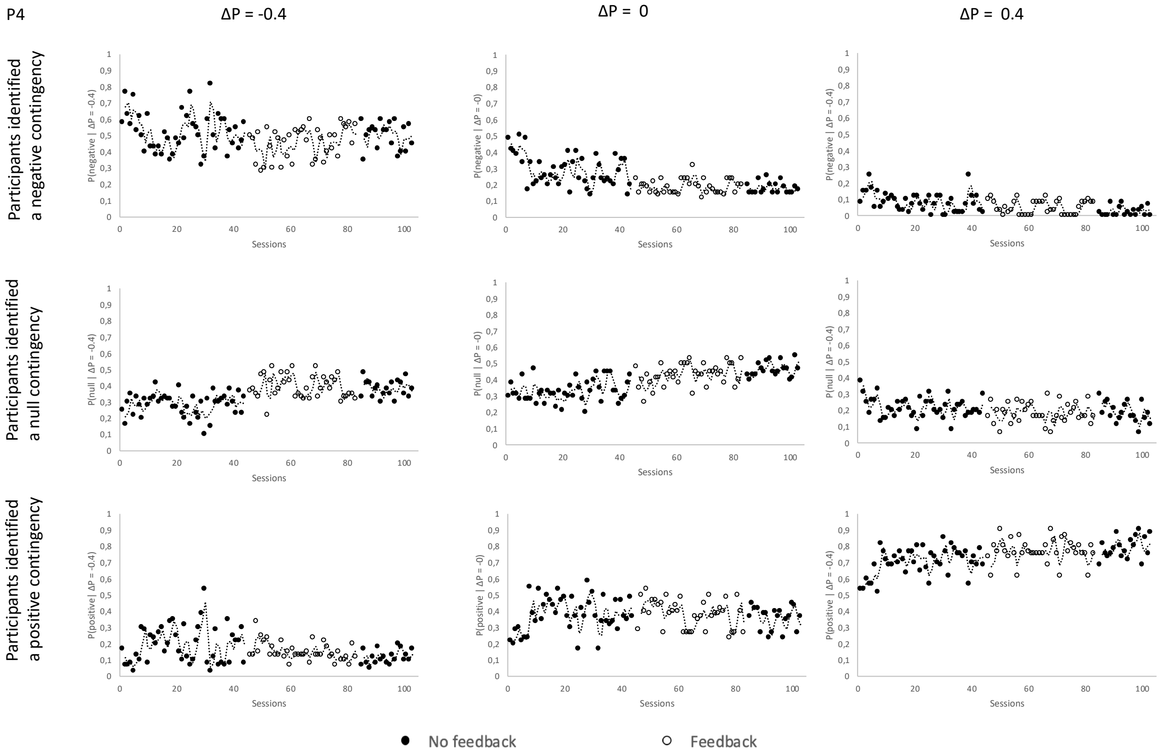


Figure 4. Top: Probability to identify a specific contingency (positive, null, or negative) as a function of the objective contingency across sessions and conditions for participant P3. Bottom: Proportion of trials for which participant P3 identified a specific contingency

(positive, null, or negative) while claiming to be sure as a function of the objective contingency across sessions and conditions. The dashed line is an exponential moving average with a window of 2.



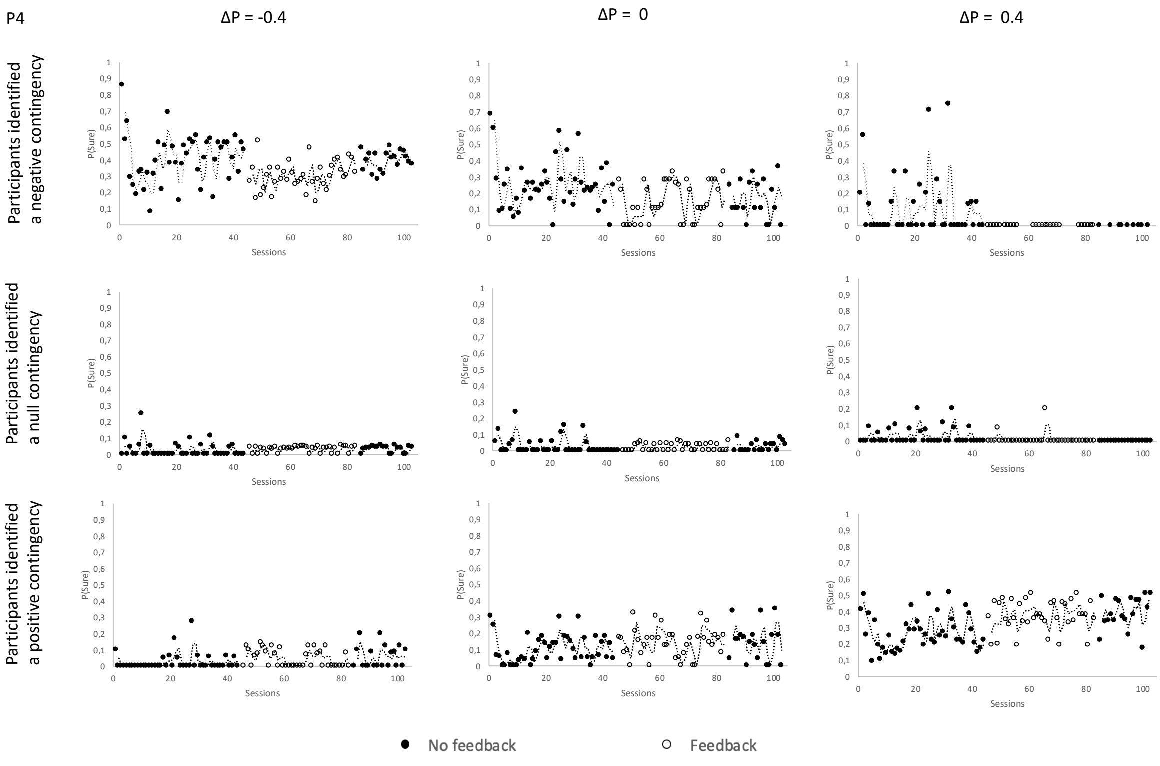


Figure 5. Top: Probability to identify a specific contingency (positive, null, or negative) as a function of the objective contingency across sessions and conditions for participant P4. Bottom: Proportion of trials for which participant P4 identified a specific contingency

(positive, null, or negative) while claiming to be sure as a function of the objective contingency across sessions and conditions. The dashed line is an exponential moving average with a window of 2.

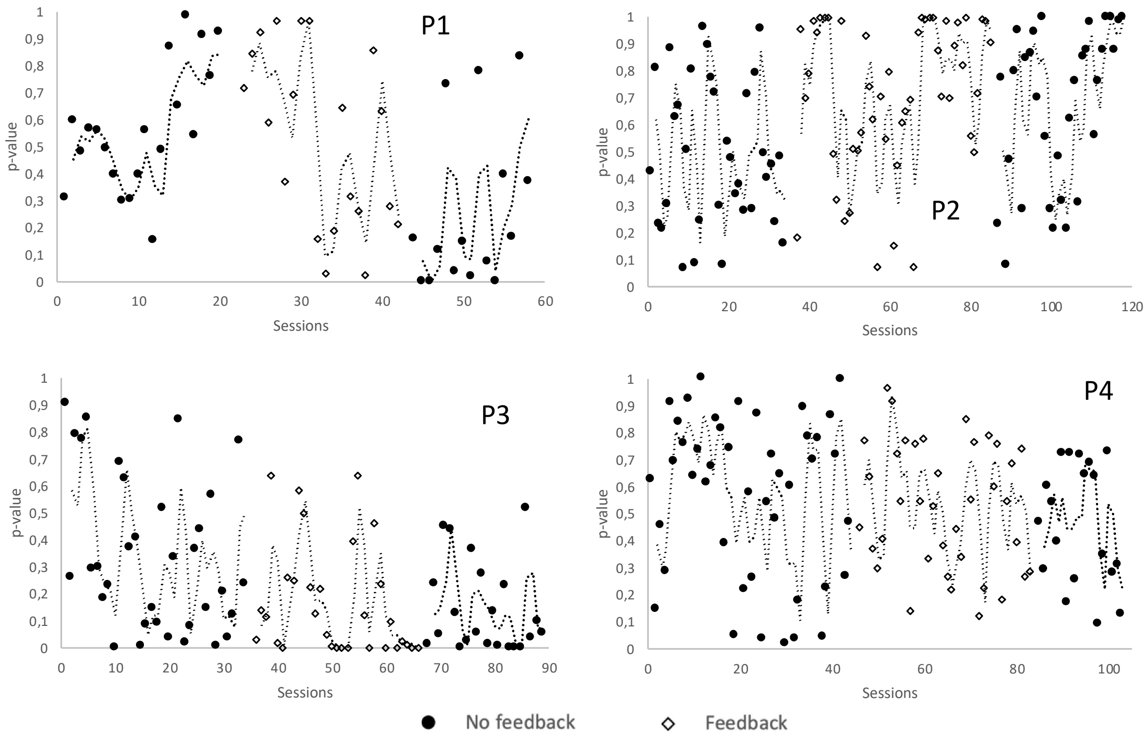


Figure 6. *P-value for each participant across sessions and conditions for the chi square measuring the fit of the SDT model to the data. The dashed line is an exponential moving average with a window of 2.*

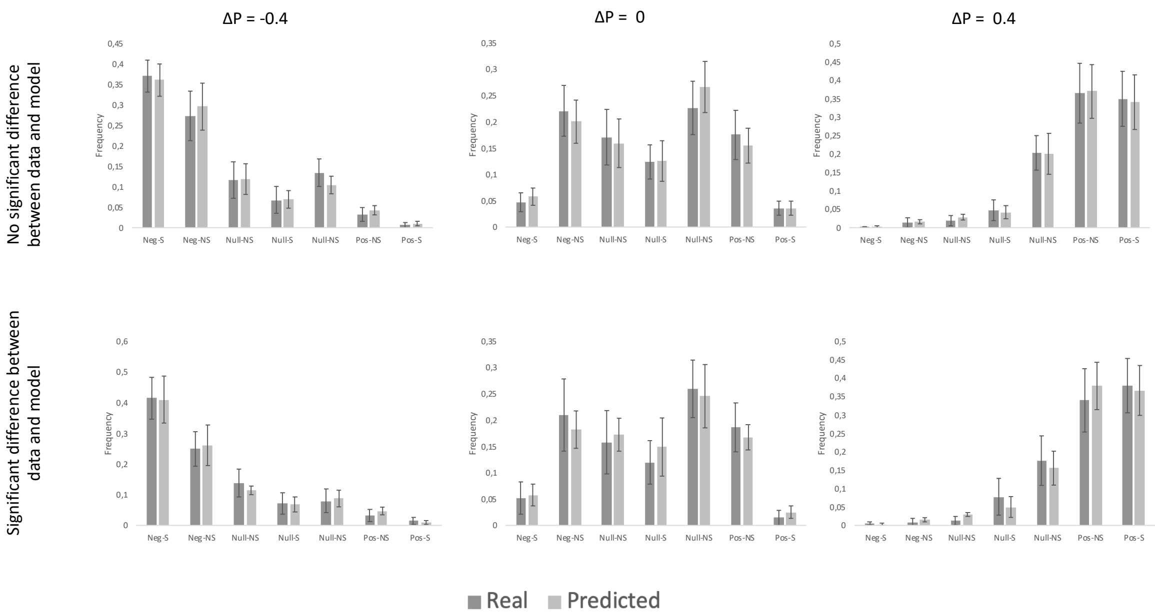


Figure 7. Probability to provide a given rating as a function of the objective contingency for participant P3 along with the prediction of the SDT model. Top: Statistical analysis did not detect any significant difference between the model and the data. Bottom: Statistical analysis detected a significant difference between the model and the data. Error-bars are 95% CI.

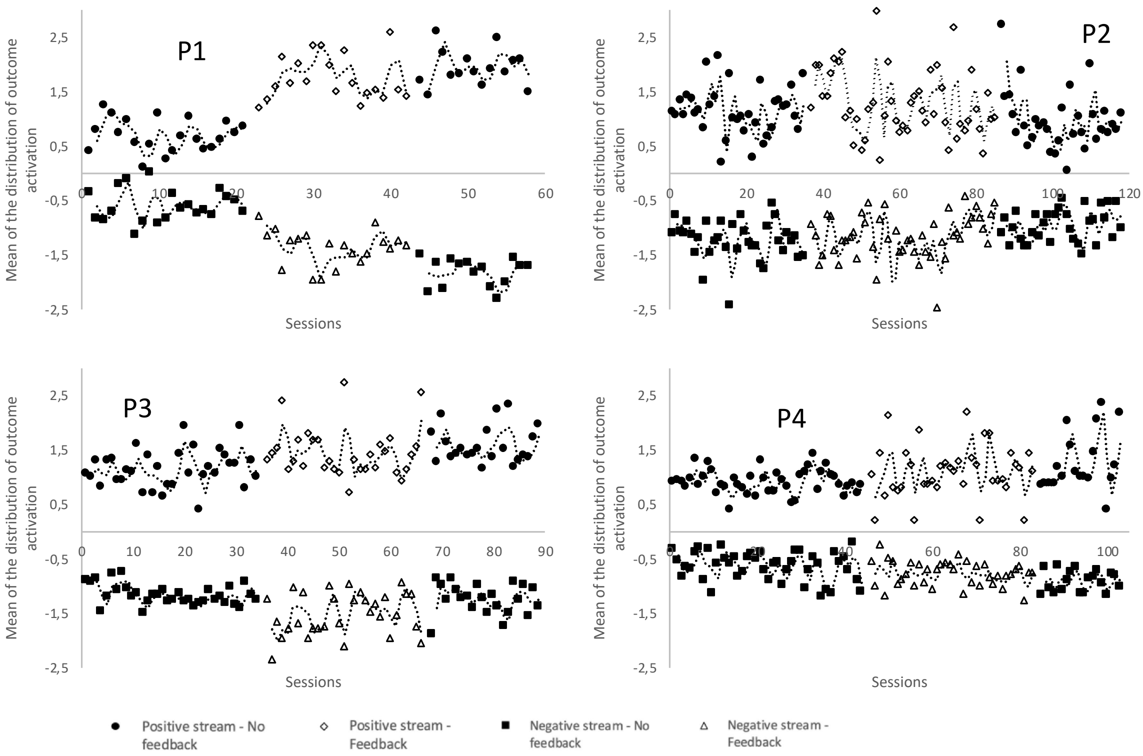


Figure 8. *Mean of the distributions of outcome activation for each participant across sessions and conditions. The mean for the null contingency is set at 0. The dashed line is an exponential moving average with a window of 2.*

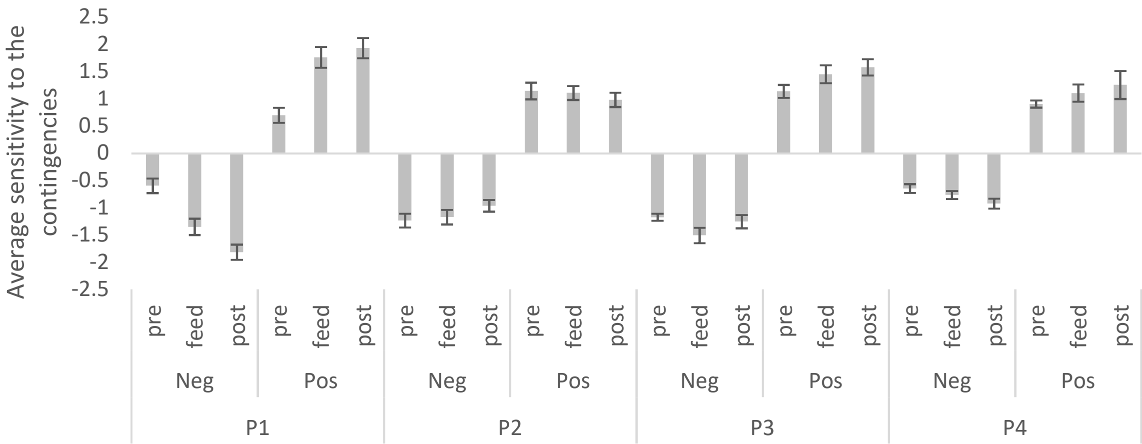


Figure 9. Average sensitivity to the positive and negative contingencies as a function of conditions for each participant. The mean for the null contingency is set at 0. Error-bars are 95% CI.

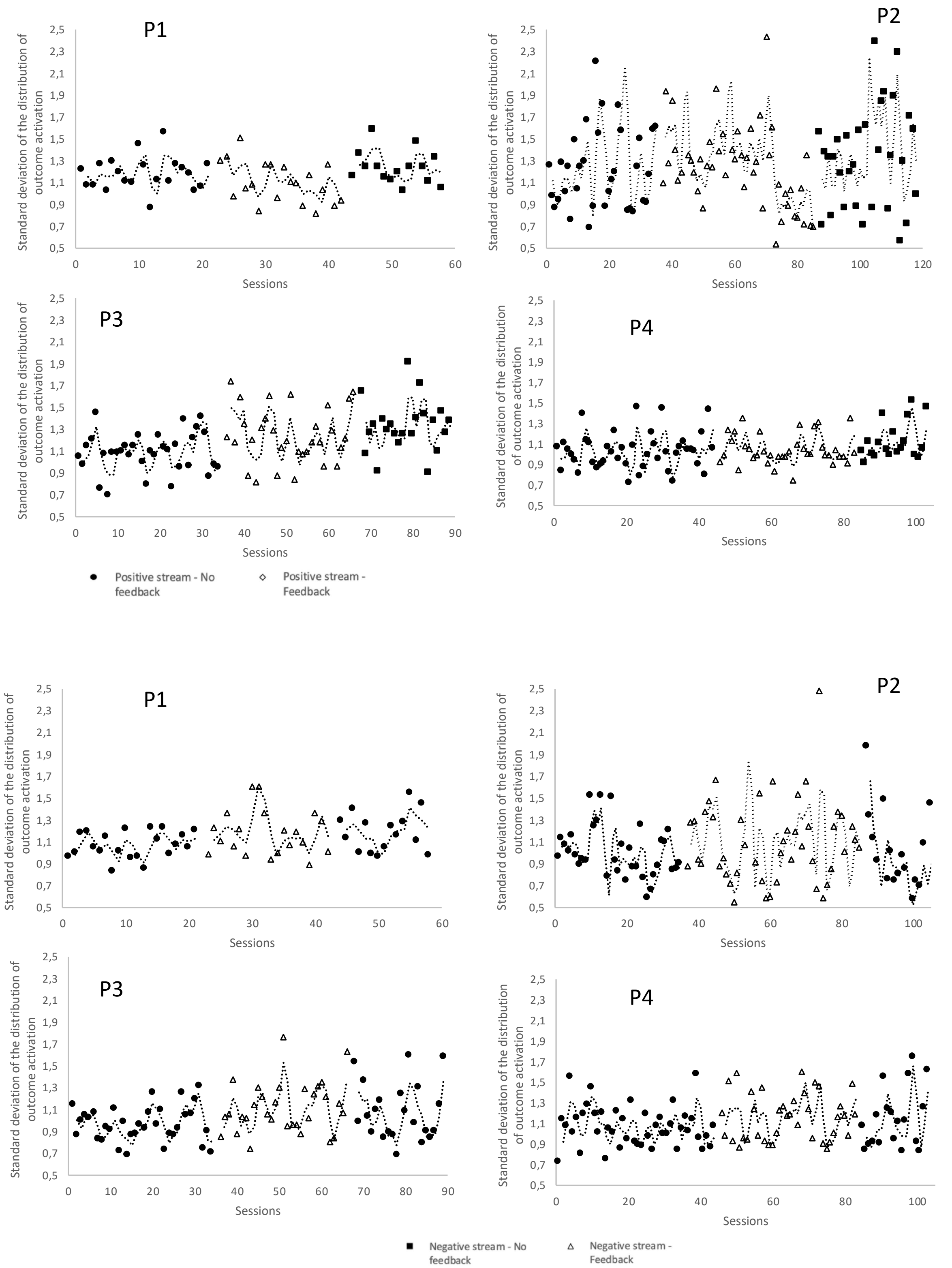


Figure 10. Standard deviation of the distributions of outcome activation for each participant across sessions and conditions. The standard deviation for the null contingency is set at 1. The dashed line is an exponential moving average with a window of 2.

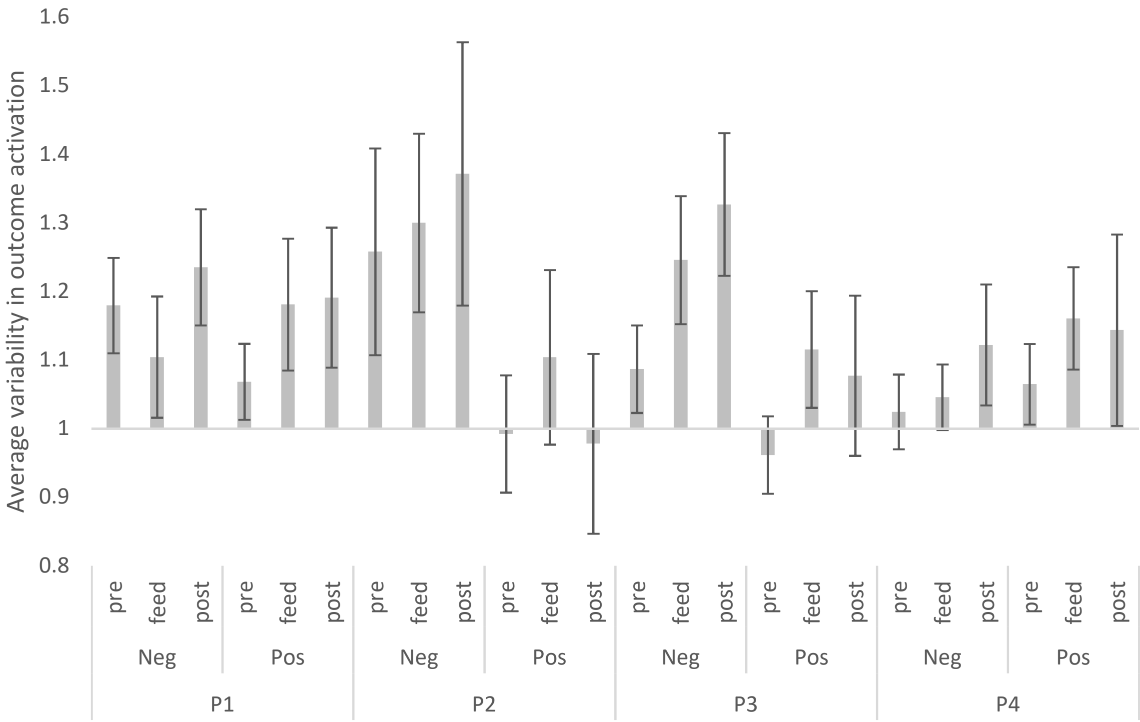


Figure 11. Average variability in outcome activation for the positive and negative contingencies as a function of condition for each participant. The standard deviation for the null contingency is set at 1. Error-bars are 95% CI.

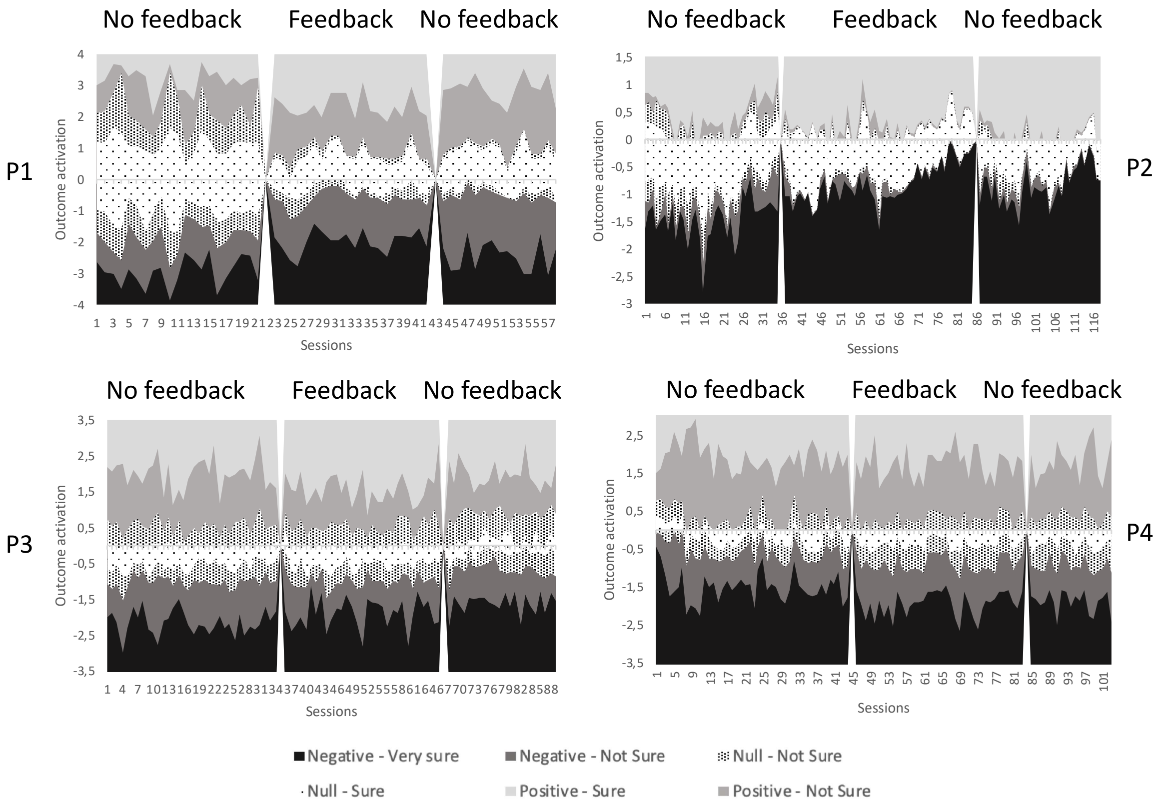


Figure 12. Decision area across sessions for each participant.