Received: 28 February 2018 | Revised: 20 March 2019 | Accepted: 20 March 2019
DOI: 10.1002/bdm.2129

RESEARCH ARTICLE

WILEY

Forward inference in risky choice: Mapping gaze and decision processes

Martin Schoemann[1] | **Michael Schulte-Mecklenbeck**[2,3] | **Frank Renkewitz**[4] | **Stefan Scherbaum**[1]

[1] Department of Psychology, Technische Universität Dresden, Dresden, Germany

[2] Department of Business Administration, University of Bern, Bern, Switzerland

[3] Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany

[4] Department of Psychology, University of Erfurt, Erfurt, Germany

Correspondence
Martin Schoemann, Department of Psychology, Technische Universität Dresden, Zellescher Weg 17, 01069 Dresden, Germany.
Email: martin.schoemann@tu-dresden.de

Funding information
Start-up grant; Department of Psychology, Technische Universität Dresden

Abstract

The study of cognitive processes is built on a close mapping between three components: overt gaze behavior, overt choice, and covert processes. To validate this overt–covert mapping in the domain of decision-making, we collected eye-movement data during decisions between risky gamble problems. Applying a forward inference paradigm, participants were instructed to use specific decision strategies to solve those gamble problems (maximizing expected values or applying different choice heuristics) during which gaze behavior was recorded. We revealed differences between overt behavior, as indicated by eye movements, and covert decision processes, instructed by the experimenter. However, our results show that the overt–covert mapping is for some eye-movement measures not as close as expected by current decision theory, and hence question reverse inference as being prone to fallacies due to a violation of its prerequisite, that is, a close overt–covert mapping. We propose a framework to rehabilitate reverse inference.

KEYWORDS

eye-tracking, expected value, heuristic, process tracing, reverse inference, risky choice

1 | INTRODUCTION

The field of judgment and decision-making (JDM) is subjected to an ongoing paradigmatic shift as it evolves its focus, methods, and approaches from an outcome-based/economic perspective toward a more process-orientated/psychological paradigm (Oppenheimer & Kelso, 2015; Schulte-Mecklenbeck et al., 2017). This process paradigm explains decision-making through perceptual, attentional, memory, and aggregation processes. Thus, process models of decision-making have become increasingly important as it is now of superordinate interest how environmental information is translated into choices, that is, the field of JDM has shifted its focus from what people choose to how they decide (Johnson & Ratcliff, 2014). However, the investigation of those processes requires methodologies that allow for correct inferences of covert cognitive processes based on observable behavior. The evaluation and expansion of such a methodology is the objective of this article.

A prerequisite for any methodology to investigate cognitive processes is a close relationship between overt observable behavior and the underlying covert cognitive processes. The measurement of eye movements is one approach that relies on that specific assumption in order to provide insights into the processes underlying a wide variety of human behaviors such as walking, driving, reading, and decision-making (Glaholt & Reingold, 2011; Hayhoe & Ballard, 2005; Orquin & Mueller Loose, 2013; Rayner, 2009). Following Poldrack (2006), the reasoning in eye-tracking studies then works as follows:

1. When a task recruits some psychological process P, eye-movement pattern EM is likely to be found;

2. in the present study, a pattern EM was found when task T was presented; it can therefore be concluded that
3. the psychological process P was recruited by task T.

This reasoning is deductively invalid because a variety of cognitive processes can be responsible for the same observable patterns of, for instance, blood oxygenation level-dependent signals or even eye movements and was denominated as the fallacy of reverse inference (Glymour & Hanson, 2016; Poldrack, 2006). In defense of reverse inference, Machery (2014) has argued that this fallacy can be overcome by the comparison of competing hypotheses and hence comparative conclusions. Indeed, this kind of model comparison is typical for eye-tracking studies of decision-making (e.g., Fiedler & Glöckner, 2012) seemingly excluding those from being subject of the reverse inference critics.

However, even the comparative conclusions drawn from model comparisons are vulnerable as they still rely on the first premise of the aforementioned reverse-inference reasoning. Accordingly, we assume a somewhat high probability that a cognitive process produces a specific eye-movement pattern—the mapping between covert cognitive processing and overt observable behavior. In fact, we do not have sufficient reasons for this assumption. Thus, the same cognitive process might produce a variety of eye-movement patterns that we would in turn—employing a reverse inference—wrongly credit to a variety of cognitive processes. Even if we test competing hypotheses—as argued by Machery (2014)—this variety of eye-movement patterns may still lead us to favor the wrong model. In Marr's (1982) terminology (cf. Anderson, 1990; Pylyshyn, 1984), we, hence, describe covert decision processes (e.g., heuristics) at the computational level where problems are specified generically. In contrast, we measure overt behaviors (e.g., eye movements) at the algorithmic level showing how exactly the computational problems are solved. Thus, the same decision process can be realized in a variety of overt behaviors, and hence, the mapping between covert cognitive processing and observable behavior is prone to great variability.

To deal with the reverse-inference problem, one prominent solution are multimethod approaches (e.g., Forstmann, Wagenmakers, Eichele, Brown, & Serences, 2011; Turner, Forstmann, Love, Palmeri, & Van Maanen, 2017; Turner, Rodriguez, Norcia, McClure, & Steyvers, 2016), using multiple paths to connect descriptions (i.e., models) and observations (i.e., data) on different levels (cf. Marr, 1982). Another, to-date less prominent solution, forward inference has been introduced by Henson (2006). Forward inference turns the direction of inference upside down by making cognitive processes explicit through instruction (Heit, 2015; Henson, 2006). By this means, the mapping between observable behavior and the cognitive process, as instructed by the experimenter, can be evaluated.

Schulte-Mecklenbeck, Kühberger, Gagl, and Hutzler (2017) recently applied a paradigm in risky choice that allows for such a forward inference. They evaluated the mapping between eye movements and different process predictions derived from distinct decision models. The derivation of process predictions was justified by the information processing perspective on decision-making, which argues that a model can be broken down into a sequence of successive cognitive operations, or elementary information processes (EIPs) such as read, compare, difference, add, product, eliminate, move, and choose (Bettman, Johnson, & Payne, 1990; Payne, Bettman, & Johnson, 1988, 1993). With the exception of read and choose, different decision models use different EIPs and/or different distributions of the respective EIPs; for example, calculating an expected value (EV) would use product whereas using the priority heuristic (PH, see below) would use compare (among other EIPs)—these differences should result in different eye-movement patterns. In the Schulte-Mecklenbeck, Kühberger, et al. (2017) paradigm, participants first learned these decision strategies (EV and PH) and were then instructed to apply them on the same set of risky gambles. These decision strategies differ strongly with respect to the predictions on information acquisition and processing that should be captured by the measurement of eye movements. Schulte-Mecklenbeck, Kühberger, et al. (2017) found misfits between the two models' predictions and participants' gaze behavior and concluded that overt behavior, indicated by eye movements, is to some extent ambiguous with respect to covert, underlying cognitive (decision) processes. They concluded that at least some studies of decision-making in that eye-movement patterns were interpreted within some (narrow) theoretical account committed the fallacy of the reverse inference (Brandstätter & Körner, 2014; Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; Pachur, Hertwig, Gigerenzer, & Brandstätter, 2013; Su et al., 2013; Zhou et al., 2016).

Before one reaches the conclusion that a long line of previous research has committed the fallacy of the reverse inference, one should carefully examine the inevitability of the conclusion drawn by Schulte-Mecklenbeck, Kühberger, et al. (2017). Its validity could be limited, for example, due to methodological issues; while a forward inference paradigm is an elegant way to evaluate the mapping between overt and covert processes, it should comply with (at least) three methodological features that Schulte-Mecklenbeck, Kühberger, et al. (2017) did not implement. Those features are related to (a) the set of used gamble problems, (b) the manipulation check, and (c) the incentives. First, the set of gamble problems should be tailored in a way that the correct application of each instructed decision strategy leads to a distinct choice. This feature eventually enables participants to apply those strategies and hence to make correct choices according to the respective decision strategy. However, making correct choices cannot be the only criterion to ascertain that participants effectively applied the instructed strategies, as it is not possible to tailor sets of risky gamble problems that allow an outcome-based strategy classification for more than two decision strategies (Bröder, 2010). Second, it is beneficial to conduct a strategy check, that is, testing whether participants effectively applied the strategy they were instructed. Third, a forward inference approach should have different incentives as participants shall not maximize their payoff by choosing between gamble problems they would like to be played out, but rather by solving gamble problems according to the instructed strategies. Hence, participants should be incentivized for making correct choices in the light of the respective strategy.

TABLE 1 Decision rules of three decision strategies instructed in the current forward inference paradigm; taken from Brandstätter et al. (2006) and Payne et al. (1993)

Strategy	Decision rule
The priority heuristic	Go through reasons in the following order: minimum outcome, probability of minimum outcome, and maximum outcome. Stop examination if the minimum outcomes differ by 1/10 (or more) of the maximum outcome; otherwise, stop examination if probabilities differ by 1/10 (or more) of the probabilities scale. Choose the gamble with the more attractive outcome (probability).
Expected value theory	Calculate the sum of all weighted possible outcomes within each option of the gamble problem using the formula $\sum P_i * O_i$, where P_i and O_i are the probability and outcome of the probability–outcome pair *i*. Choose the option with the highest weighted sum.
The minimax heuristic	Choose the option with the highest minimum outcome.

In this study, we propose a forward inference paradigm that addresses the above outlined shortcomings. Thereby, we scrutinize previous findings from Schulte-Mecklenbeck, Kühberger, et al. (2017) and reassess their conclusion that (at least some) previous studies of decision-making, in which eye-movement patterns were interpreted within some (narrow) theoretical account, committed the fallacy of the reverse inference due to a violation of its prerequisite.

We applied a risky choice paradigm with three instructed decision strategies—the PH (Brandstätter, Gigerenzer, & Hertwig, 2006), EV (Hacking, 1984), and the minimax heuristic (MM; Coombs, Dawes, & Tversky, 1970, p. 141). Detailed descriptions of these strategies can be found in Table 1. We instructed participants to apply these strategies on the same set of gamble problems. All gamble problems could be solved applying either strategy. Each of the strategies (PH, EV, and MM) make fundamentally different process predictions—hence they offer themselves as ideal test objects for our question. Whether these strategies reflect actual choice behavior is of lesser interest for this article. After each choice in a gamble problem, we implemented a strategy check, that is, participants had to indicate properties of the previous gamble problem of which they should only be aware if they applied the demanded strategy. Participants were rewarded if they chose according to the strategy and answered correctly in the strategy check. Hence, the different set of gamble problems, the strategy check and the reward for both, correct choice and successful strategy check were the major improvements of our forward inference paradigm in comparison with Schulte-Mecklenbeck, Kühberger, et al. (2017). Additionally, we applied a within-subjects experimental design rather than a between-subjects design that has been used in Schulte-Mecklenbeck, Kühberger, et al. (2017) for the comparison between the EV and PH condition. We also added the less complex MM condition in order to introduce variance in strategy complexity. On the level of stimulus presentation, we also switched from an equidistant 2 × 4 matrix format to an equidistant ellipsoid format (see Section 2.4). We expected that these improvements would maximize the probability that participants applied the instructed strategies, which in turn would enhance a valid evaluation of the mapping between overt and covert processes.

2 | METHOD

2.1 | Ethics statement

The study was performed in accordance with the guidelines of the Declaration of Helsinki and of the German Psychological Society. An ethical approval was not required because the study did not involve any risk or discomfort for the participants. All participants were informed about the purpose and the procedure of the study and gave written informed consent prior to the experiment. All data were analyzed anonymously.

2.2 | Participants

Forty participants (45.16% female, mean age = 23.19, *SD* = 2.95) completed the experiment conducted at the Technische Universität Dresden, Germany. The experiment lasted 75 minutes. Participants were uniformly recruited through the department's data-base system, which is based on ORSEE (Greiner, 2004). Participants received either class credit or €5.00 show-up fee as well as up to €5.04 bonus depending on individual performance. All participants had normal or corrected-to-normal vision. Nine participants were excluded from subsequent analyses due to major technical issues concerning the eye-tracking procedure such as finding the corneal reflection point and hence run a proper calibration (Participants 1, 2, 15, 26, and 28), participants' problems applying the instructed strategies (14, 16, 18),[1] and data loss (34). If we included the participants with problems of applying the instructed strategies, the results did not show qualitative differences.

2.3 | Apparatus

Eye-movements were recorded with an EyeLink 1000 desk-mounted eye tracker (SR Research, Ontario, Canada), which has a reported average accuracy between 0.25° and 0.50° of visual angle and root mean square (RMS) resolution of 0.01° (www.sr-research.com). Participants' right eye movements were tracked using the combined pupil and corneal reflection setting at a sampling rate of 250 Hz. A chin rest was used to minimize head movements and to hold the distance between the eye and the monitor (BenQ E910 LCD) at 82 cm. The screen had a diagonal of 19 inches and a resolution of 1,280 × 1,024 pixels; the refresh rate was 60 Hz. Stimuli were presented in Courier New font (30 point = 1 cm); a single letter corresponded to about 0.70° of visual angle. Preceding each of the three experimental blocks, we ran nine-point grid calibration

[1]Even though those participants solved the warm-up gambles (see Procedure) correctly, they showed high error rates that had already been noticed by the experimenter during the experiment (see error protocol available online). The corresponding occurrence of those two indicators led to the exclusion eventually.

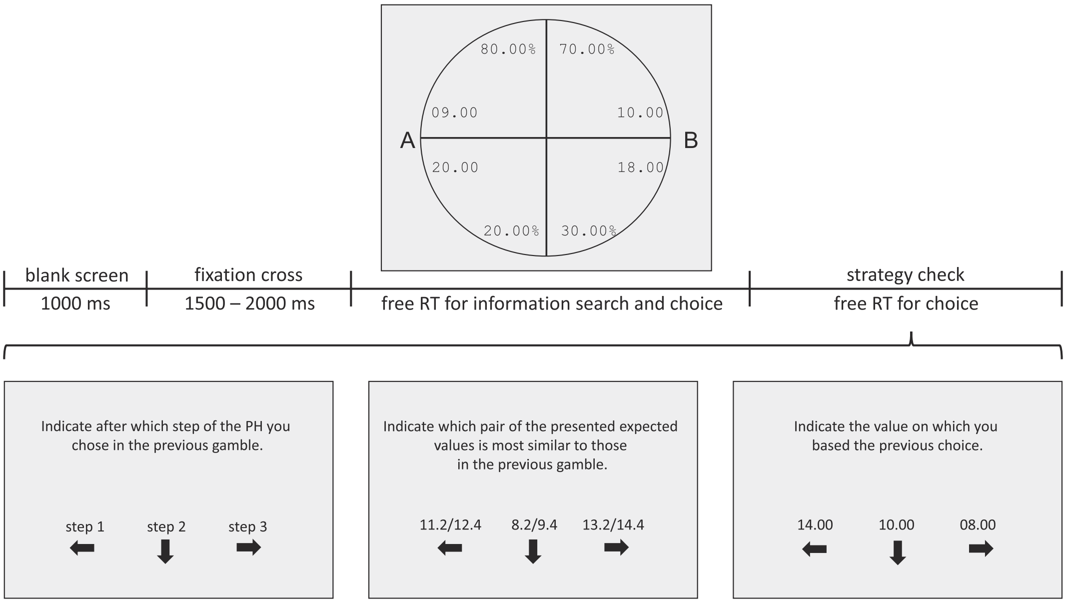

FIGURE 1 Within trial procedure in the experiment (top panel) and example of the strategy check for each strategy: priority heuristic (PH), expected value, and minimax heuristic (bottom panel from left to right). Note: the presentation time of the fixation cross was randomized with values drawn from a uniform distribution in the range from 1,500 to 2,000 ms. Choices of both during the gamble and the strategy check were indicated by pressing the adequate direction from the arrow keys on the keyboard.

(using a grid of three horizontal positions × three vertical positions) and drift correction as well as validation of both settings and potential re-calibration; we obtained an average accuracy of 0.33° (*SD* = 0.10°) of visual angle.

2.4 | Material

In each risky gamble problem, participants had a choice between two options (A and B). Each option consisted of two outcome-probability (O-P) pairs. All outcomes (O) and probabilities (P) were presented in an ellipsoid display format (see Figure 1, top panel) in which all pieces of information (i.e., outcomes and probabilities) are present at equal distance from an initial fixation point (Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011). We constructed a set of 21 gamble problems (see Appendix A). The mean EV for Option A was €8.95 (*SD* = €3.4). The mean EV for Option B was €8.40 (*SD* = €3.6). The mean absolute differences in EVs between the options was €1.82 (*SD* = €0.98). Each gamble could be solved unambiguously using either strategy. Within these 21 gamble problems, we created three subsets of seven gamble problems in which the PH terminated after the first, second, and third step, respectively (one-, two-, or three-reason gamble problems, see Brandstätter & Körner, 2014). In one-reason gamble problems, the decision should be based on the comparison of minimum outcomes only. The minimum outcomes are given by the smallest outcome in either option. In two-reason gamble problems, the decision should be based on the comparison of minimum probabilities because throughout the previous comparison of the minimum outcomes, none of the options was clearly dominant. The minimum probabilities are given by the respective adjacent probabilities to the minimum outcomes. In three-reason gamble problems, the former logic extends to the comparison of the maximum outcomes. The maximum outcomes are given by the highest outcome in either option.

After each risky gamble problem, participants had to indicate whether they had applied the correct strategy—the strategy check (see Figure 1, bottom panel). In the PH condition, we asked after which step the PH was terminated. In the EV condition, participants were confronted with three pairs of EVs, and they had to indicate which of those pairs describes the preceding gamble. In the MM condition, participants were confronted with three outcomes, and they had to indicated which of those outcomes was the outcome they were instructed to base their decision on.[2]

We provided feedback after each strategy check indicating whether participants made the correct choice as well as answered the strategy check correctly. Therefore, we used a 2-s increasing tone sequence for positive feedback, and a 1-s decreasing tone sequence for negative feedback (available online).

2.5 | Procedure

Participants were first given information about the eye-tracking setup and a general description of the gamble problems. Then, all participants worked through three experimental blocks with the order of the blocks being counterbalanced between participants. Each block started with a detailed instruction of the respective strategy, consisting of an explanation of each step necessary to follow the

[2]All items were randomly computed online; we provide the Matlab script for this calculation online.

strategy as well as a working example (available online). After the instruction, participants and instructors worked through three warm-up gambles to ensure strategy comprehension. To that end, during the three warm-up gamble problems, participants were urged to verbalize their cognition and instructors were free to correct participants or answer any questions concerning the respective strategy. After the warm-up gambles, the eye-tracking system was calibrated. Participants were then presented with the 21 risky gamble problems in within-block randomized order and asked to choose the correct gamble option (A, B) given the instructed strategy. Each risky gamble problem was presented three times, and hence, participants solved 63 gamble problems throughout the experiment. Due to the requirements of a planned scanpath analysis, gambles were presented in a fixed setup in each of the three respective occurrences across all participants, that is, all participants were presented with the same stimuli three times. The results of the scanpath analysis will be reported elsewhere.

After each of the 21 gamble problems, the strategy check was applied (see Section 2.4). When participants correctly chose and succeeded in the strategy check, they earned a bonus of €0.08, indicated by a specific sound (see Section 2.4). Given a wrong response (choice or strategy), no bonus was awarded, indicated by a different sound. Participants could earn a bonus of up to €5.04 (3 * 21 gambles * €0.08). At the end of each block, participants were informed about the bonus they had earned so far.

2.6 | Data preparation and processing

Eye-movement data were analyzed with a customized Matlab script (available online). Saccade and fixation events were defined on-line by the host[3] of the eye-tracking system and read out by the script if a fixation landed on one of the pre-defined, non-overlapping areas of interest (AOIs). Eight AOIs with a size of 150 × 80 pixels were defined with a visual angle of 3.70° × 1.96° and the numeric values in the center. The minimal horizontal distance between the AIOs of the outcomes of the two gambles (e.g., 80% and 70%, or 20% and 30%, see Figure 1) was 150 pixels (3.70°); the minimal vertical distance between the outcomes as well as outcomes and probabilities within gambles (e.g., 9 and 20, or 80% and 9, see Figure 1) was 120 pixels (2.96°), respectively. Fixation durations shorter than 50 ms were removed from the analysis. Consecutive fixations on the same AOI were concatenated and hence denominated as dwells.

On a trial level, all data were included regardless of whether the wrong option was chosen or the strategy check was answered incorrectly. We did so, because participants might have applied the instructed strategies erroneously; instead, we excluded participants with high error rates proposing that those participants have not been able to apply the instructed strategies (see Section 2.2).

Data preprocessing and aggregation was performed in Matlab 2015a (the Mathworks Inc.). Statistical analyses were performed in Matlab, and R (R Core Team, 2018). In order to ease comparison of results between the current and the Schulte-Mecklenbeck, Kühberger, et al. (2017) study, we applied the same statistical analyses using the same set of R packages. Additionally, we reanalyzed the Schulte-Mecklenbeck data using only gambles that involved gains and report these results were appropriate in footnotes.

2.7 | Predictions

Our analysis will focus on participants' choices as a precondition for the successful application of the instructed strategy and on the following process measures: (a) dwell patterns, (b) dwell frequencies, and (c) dwell times. For each of these dependent measures, we will derive a separate set of predictions. Table 2 provides a detailed set of predictions (and results) for dwell frequencies (also specifying expected differences for one-, two-, and three-reason gamble problems).

2.7.1 | Choices

In the MM condition, participants were instructed to base their decision on the mere comparison of the options' minimum outcome (O^{min}). In the PH and EV conditions, participants were supposed to execute both comparisons and calculations. PH demands more comparison and less calculation, whereas EV demands less comparison and more calculation. Speaking in terms of EIPs, all strategies demand read and choose, but EV and PH additionally demand both compare and product, whereas MM merely demands compare. Hence, MM can be interpreted as less complex (or difficult to apply) than PH and EV, and therefore we expect participants' choices to be most consistent with the strategy they were instructed to use in the MM condition. Because product (as necessary for EV) is more prone to errors than compare (mainly applied in PH) as well as based on recent results (see Schulte-Mecklenbeck, Kühberger, et al., 2017), we expect choice consistency in PH condition being higher than in EV.

2.7.2 | Dwell patterns

In the EV condition, we instructed participants to base their decision on the calculation and comparison of EVs, for example, to multiply outcome (O) and probability (P). Hence, we expect frequent within-option transitions O → P or P → O. In the PH condition, we instructed participants to base their decisions mostly on pairwise comparisons, for example, to compare minimum outcomes. Hence, we expect frequent O → O or P → P transitions between options, depending on the necessary number of comparisons (one-, two-, or three-reason gambles). In the MM condition, we instructed participants to base their decisions on the mere comparison of minimum outcomes. Hence, we expect frequent O → O transitions between options (comparable with one-reason gambles in PH, see Section 2.4).

[3]The EyeLink software applies a velocity-based algorithm combined with acceleration criteria to detect saccade onsets and offsets. We used a *cognitive configuration* combining velocity, acceleration, motion and pursuit thresholds of 30°/s, 8,000°/s^2, 0.15°, and 60°, respectively. The motion threshold is used to ensure that the eye has moved sufficiently before a saccade is detected. The pursuit threshold is used to limit the amount that the velocity threshold can be raised by the average velocity over the last 40 ms in order to detect long and smooth eye movements. Hence, the cognitive configuration is conservative, less sensitive to noise and ignores most saccades smaller than 0.6°.

TABLE 2 Comparisons between outcomes and probabilities with predictions of dwell frequencies separately for EV, PH, and MM

	Predictions			Empirical data		
Comparison	EV	PH	MM	EV	PH	MM
O vs. P	$O_{r=1} = P_{r=1}$	$O_{r=1} > P_{r=1}$	$O_{r=1} > P_{r=1}$	56.6 > 43.4	86.7 > 13.3	98.5 > 1.5
	$O_{r=3} = P_{r=3}$	$O_{r=3} > P_{r=3}$	$O_{r=3} > P_{r=3}$	55.6 > 44.4	69.7 > 30.3	98.0 > 2.0
O^{max} vs. O^{min}	$O^{max}_{r=1} = O^{min}_{r=1}$	$O^{max}_{r=1} < O^{min}_{r=1}$	$O^{max}_{r=1} < O^{min}_{r=1}$	28.3 = 28.3	43.1 = 43.6	48.1 < 50.4
	$O^{max}_{r=2} = O^{min}_{r=2}$	$O^{max}_{r=2} < O^{min}_{r=2}$	$O^{max}_{r=2} < O^{min}_{r=2}$	27.0 = 29.1	33.2 = 31.7	47.8 = 49.7
P^{max} vs. P^{min}	$P^{max}_{r=2} = P^{min}_{r=2}$	$P^{max}_{r=2} < P^{min}_{r=2}$	$P^{max}_{r=2} = P^{min}_{r=2}$	21.3 = 22.6	17.7 = 17.4	1.3 = 1.1
	$P^{max}_{r=3} = P^{min}_{r=3}$	$P^{max}_{r=3} < P^{min}_{r=3}$	$P^{max}_{r=3} = P^{min}_{r=3}$	22.3 = 22.1	15.7 = 14.6	0.9 = 1.1
$O^{max}_{r=1}$, $O^{max}_{r=2}$ vs.$O^{max}_{r=3}$	$O^{max}_{r=1} = O^{max}_{r=3}$	$O^{max}_{r=1} < O^{max}_{r=3}$	$O^{max}_{r=1} = O^{max}_{r=3}$	28.3 = 27.7	43.1 = 36.7	48.1 = 48.7
	$O^{max}_{r=2} = O^{max}_{r=3}$	$O^{max}_{r=2} < O^{max}_{r=3}$	$O^{max}_{r=2} = O^{max}_{r=3}$	27.0 = 27.7	33.2 < 36.7	47.8 = 48.7
$P^{min}_{r=1}$ vs.$P^{min}_{r=2}$, $P^{min}_{r=3}$	$P^{min}_{r=1} = P^{min}_{r=2}$	$P^{min}_{r=1} < P^{min}_{r=2}$	$P^{min}_{r=1} = P^{min}_{r=2}$	21.9 = 22.6	6.9 < 17.4	1.0 = 1.1
	$P^{min}_{r=1} = P^{min}_{r=3}$	$P^{min}_{r=1} < P^{min}_{r=3}$	$P^{min}_{r=1} = P^{min}_{r=3}$	21.9 = 22.1	6.9 < 14.6	1.0 = 1.1

Note. Empirical findings are shown as percentages of dwells per gamble problem (e.g., O vs. P). All differences are significant on a Bonferoni corrected alpha-level ($p < .005$) as given by a paired sample t test; when models predicted no difference, test was conducted two-tailed; when models predicted difference, test was conducted one-tailed in the predicted direction of the difference.

Abbreviations: EV, expected value; MM, minimax heuristic; O, outcomes; P, probabilities; PH, priority heuristic.

2.7.3 | Dwell frequencies

In the EV condition, we expect no difference in the number of dwells on O and P. In the PH condition, the number of dwells O and P is dependent on the number of reasons. In one-reason gambles, choice is based on outcomes only (maximum outcomes to calculate aspiration level and comparing minimum outcomes). In two-reason gamble problems, both probabilities of the minimum outcomes must be considered additionally. In three-reason gamble problems, a reconsideration of the maximum outcomes can be expected. Thus, in the PH condition, we expect more dwells Os than Ps, with the ratio depending on the number of reasons. In the MM condition, we expect more dwells Os than Ps. Based on recent research, we speculated that the dwell frequency is lowest in MM and does not differ in PH and EV (see Schulte-Mecklenbeck, Kühberger, et al., 2017).

2.7.4 | Dwell times

The dwell time—that is, the duration between two saccades in that the eye rests on a specific area of interest—is taken to be indicative of attention (Rayner, 2009). However, it may also indicate consumption of cognitive resources. For instance, if it is more effortful to multiply an outcome by a probability (as necessary for EV) than to compare the size of two outcomes (as necessary for MM and mainly applied in PH), the average dwell time should be longer for product than for compare. Previous research suggests that dwell time is associated with the complexity of the process executed (Velichkovsky, 1999). Hence, EV, which involves both product and sum, can be expected to result in longer dwells than PH and MM, which require only compare, despites the calculation of the aspiration level (product) in PH. Therefore, we expect longer dwells in PH than MM.

2.8 | Data statement

All primary data and analysis scripts as well as some materials are available and can be accessed at https://osf.io/q3ybt/.

3 | RESULTS

3.1 | Choices

First, we evaluated whether participants' choices were consistent with the strategy they were instructed to use: EV, PH, or MM. All strategies discriminated between gamble options in all gamble problems (see Appendix) and, hence, predicted choices distinctly. Participants' choice behavior showed that strategy instructions were effective in guiding participants to the predicted choices. Participants instructed to use EV chose the option with the higher EV in 88% of the gamble problems; participants instructed to use the PH-made corresponding choices in 88% of the gamble problems; participants instructed to use MM chose the option with the higher minimum outcome in 98% of the gamble problems. When taking into account whether the strategy check was also answered correctly, results showed a similar pattern (EV – 81%, PH – 82%, and MM – 97%) with lower overall accuracy.[4]

We also inspected participants' choice behavior separately for one-, two-, and three-reason gambles: in one-reason gamble problems, participants chose the option with the highest minimum outcome in 95%; in two-reason gamble problems, participants chose the option with highest probability of the minimum outcomes in 81%; in three-

[4]Descriptive results were supported statistically by generalized multilevel regression analysis at the trial level using a binomial link function and including participants as random intercept and condition (EV, PH, and MM) as a fixed effect (see complete analyses script online).

reason gamble problems, participants chose the option with highest maximum outcomes in 88%. Taking correct strategy checks into account, a similar pattern emerged (one-reason PH – 90%, two-reason PH – 74%, and three-reason – 81%).

Therefore, we successfully instructed participants to use a specific decision strategy during risky choice and outperformed instruction of strategy used by Schulte-Mecklenbeck, Kühberger, et al. (2017) who reported 62% and 80% choices being consistent with EV and PH, respectively. However, due to our additional strategy check, we can assume that participants' actually applied the instructed strategies, which justifies to evaluate top-down effects on different process measures such as dwell pattern, dwell frequency, and dwell time.

3.2 | Dwell patterns

In order to evaluate the overall pattern of information acquisition as indicated by dwell pattern, we determined the ratio of within- to between-option transitions by applying Böckenholt's Search Metric (SM) calculation (Böckenholt & Hynan, 1994) as follows:

$$SM = \frac{\sqrt{N}^{*}\left(\left(\frac{OA}{N}\right)(n_{opt} - n_{att}) - (A - O)\right)}{\sqrt{O^2(A-1) + A^2(O-1)}},$$

where O indicates the number of gamble options (two in our experiment), A indicates the number of attributes (four in our experiment), N indicates the total number of transitions, n_{opt} indicates the number of option-wise transitions, and n_{att} indicates the number of attribute-wise transitions. By those variables, the SM incorporates the specific stimulus setup of the gamble problem. The SM distinguishes within-option transitions (i.e., successive transitions of information within the same gamble option but between attribute dimensions, n_{opt}) from between-option transitions (i.e., successive transitions of information within the same attribute dimension but between gamble options, n_{att}). We replaced the absolute occurrences of transitions (N, n_{opt}, n_{att}) with their proportions to capture the index's sensitivity to large Ns (see Pachur et al., 2013). An SM <0 indicates a predominance of between-option transitions, and an SM >0 indicates a predominance of within-option transitions.

The evaluation of participants' overall dwell patterns showed the predicted pattern (see Figure 2). Participants' application of EV led on average to the highest, positive index, SM_{EV} = 6.10 (SD = 1.66), indicating a predominance of within-option transitions; the application of PH and MM resulted on average in a negative index, SM_{PH} = −9.19 (1.77), SM_{MM} = −11.93 (2.51), indicating a predominance of between-option transitions.

Descriptive results were supported statistically by multilevel regression analysis with participants and gamble problem as random intercepts and condition (EV, PH, and MM) as a fixed effect. The SM indeed was higher in EV than in both PH, b = 15.29, $CI_{95\%}$ = (14.98 to −15.60),[5] and MM, b = 18.02, $CI_{95\%}$ = (17.71–18.33); the difference between PH and MM was much smaller, b = 2.73, $CI_{95\%}$ = (2.42–3.04). Additionally, a similar multilevel regression analysis on a PH-only subset with participants and gamble problem as random intercepts and reasons (one reason, two reasons, and three reasons) as a fixed effect revealed that the SM was similar between one-reason and three-reason gambles, b = −0.39, $CI_{95\%}$ = (−0.81 to 0.04), but differed between one-reason and two-reason gambles, b = −1.57, $CI_{95\%}$ =) −2.00 to −1.14) as well as two-reason and three reason gambles, b = 1.18, $CI_{95\%}$ = (0.76–1.61), respectively.[6]

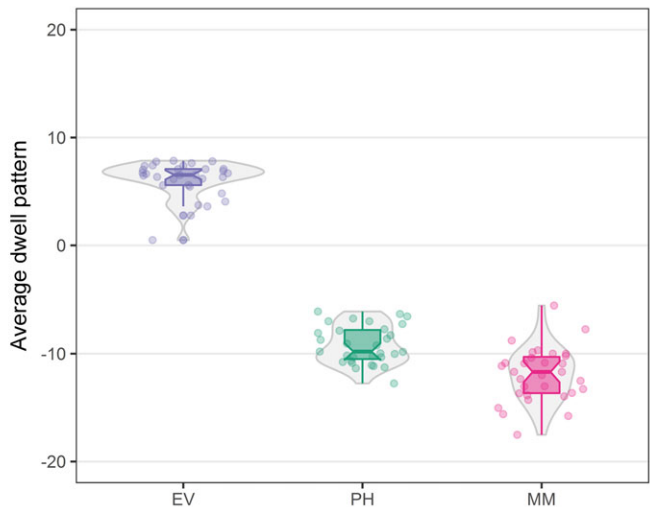

FIGURE 2 Average dwell pattern as indicated by the search metric depicted as a box plot for each of the three strategies, with superimposed averaged raw data for each participant (jittered) and a probability density function. [Colour figure can be viewed at wileyonlinelibrary.com]

In sum, as predicted, we found that participants' dwell patterns differed between all strategies and thus confirmed the finding of Schulte-Mecklenbeck, Kühberger, et al. (2017) concerning dwell patterns produced by PH and EV.

3.3 | Dwell frequencies

In order to analyze dwell frequencies, we first examined the three strategies on an aggregate level and then turned to a more fine-grained analysis of information acquisition and processing behavior.

The average number of dwells was highest when participants applied EV, Fix_{EV} = 66.78 (SD = 24.28), indicating that a comparably large number of dwells was needed to follow this strategy (see Figure 3). Participants required less dwells when following PH and MM, Fix_{PH} = 31.01 (11.31), Fix_{MM} = 8.67 (2.30), respectively.

Descriptive results were supported statistically by multilevel regression analysis with participants and gamble problem as random intercepts and condition (EV, PH, and MM) as a fixed effect. The average number of dwells index indeed was higher in EV than in both PH,

[5]Schulte-Mecklenbeck data: EV-PH, b = 8.47, $CI_{95\%}$ = (7.94–9.00).

[6]Schulte-Mecklenbeck data: one-reason–three-reason gambles, b = 0.18, $CI_{95\%}$ = (−1.21 to 1.56); one-reason–two-reason gambles, b = 0.66, $CI_{95\%}$ = (−0.73 to 2.04); three-reason–two-reason gambles, b = 0.48, $CI_{95\%}$ = (−0.90 to 1.87).

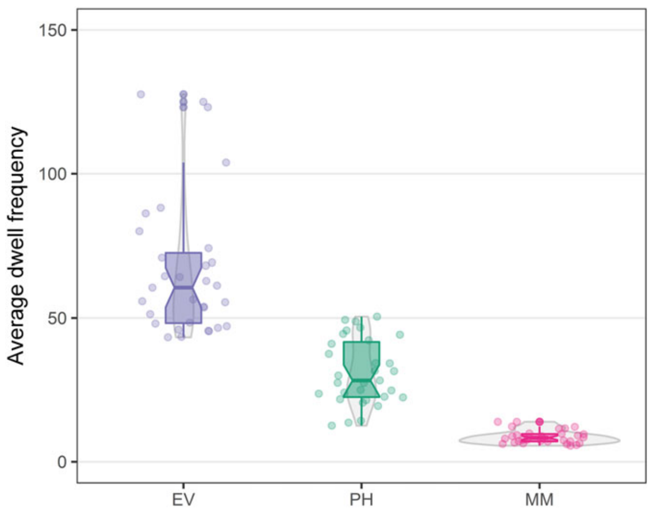

FIGURE 3 Average dwell frequency as depicted as a box plot for each of the three strategies, with superimposed averaged raw data for each participant (jittered) and a probability density function. [Colour figure can be viewed at wileyonlinelibrary.com]

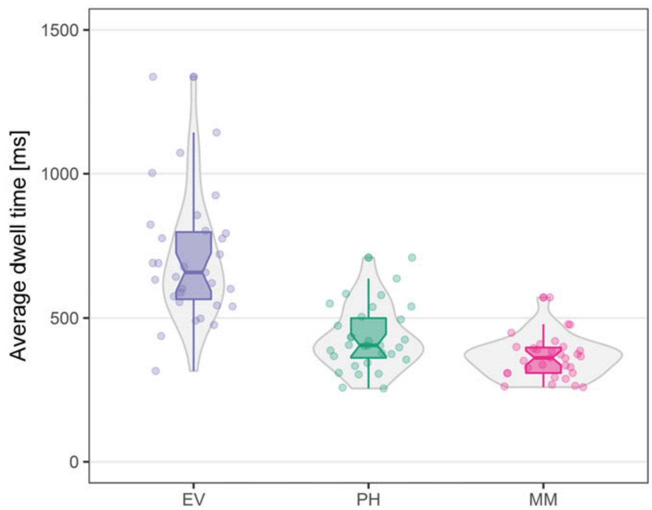

FIGURE 4 Average dwell time (in milliseconds) depicted as a box plot for each of the three strategies, with superimposed averaged raw data for each participant (jittered) and a probability density function. [Colour figure can be viewed at wileyonlinelibrary.com]

b = 35.77, $CI_{95\%}$ = (33.30–38.23),[7] and MM, b = 58.23, $CI_{95\%}$ = (55.65–60.58); the average number of dwells was also higher in PH than MM, b = 22.35, $CI_{95\%}$ = (19.88–24.81).[8]

A more fine-grained analysis of dwell frequencies focused on specific predictions on the relative number of dwells on outcomes and probabilities separately (see Johnson, Schulte-Mecklenbeck, & Willemsen, 2008; Pachur et al., 2013; Schulte-Mecklenbeck, Kühberger, et al., 2017). This analysis does not only discriminate between outcomes and probabilities, but also takes the relative size of those properties into account by distinguishing between the minimum outcome, the probability of the minimum outcome, the maximum outcome, and the probability of the maximum outcome. Table 2 presents the 10 predictions derived for the three instructed strategies as well as the corresponding empirical findings from our experiment. The first line, for example, shows predictions for outcomes and probabilities (O vs. P). For one-reason gambles (r = 1), EV predicts equal numbers of dwells on outcomes and probabilities, as indicated by $O_{r\,=\,1} = P_{r\,=\,1}$. In contrast, PH and MM predict more dwells on outcomes than on probabilities, as indicated by $O_{r\,=\,1} > P_{r\,=\,1}$. Concerning the empirical findings, we report the percentages of dwell on outcomes and probabilities separately for the three strategies.

The 10 predictions in Table 2 sketch a distinct pattern for each instructed strategy: (a) EV always predicts an equal distribution of dwell frequencies, irrespective of the properties of information or the number of reasons. (b) PH predicts more dwells on outcomes than probabilities and fewer dwells on the maximum outcome than the minimum one. (c) MM also predicts more dwells on outcomes than probabilities and fewer dwells on the maximum outcome than the minimum one. Additionally, MM predicts equal distributions irrespectively of the properties of information or the number of reasons when the minimum outcome is not involved. Table 2 shows that, overall, outcomes were fixated more frequently than probabilities, supporting the findings of Schulte-Mecklenbeck, Kühberger, et al. (2017); see also Pachur et al., 2013), which is from a theoretical point of view solely for EV a surprising finding. In addition, across all strategies, we found equal distributions between large and small outcomes as well as probabilities, which deviates from the findings of Schulte-Mecklenbeck, Kühberger, et al. (2017), who consistently found more dwell on large outcomes and probabilities, respectively.

3.4 | Dwell times

The average dwell time was 705.2 ms (SD = 219.0 ms) in EV, 429.4 ms (109.6 ms) in PH, and 362.5 ms (70.9 ms) in MM (see Figure 4). Descriptive results were supported statistically by multilevel regression analysis with participants and gamble problem as random intercepts and condition (EV, PH, and MM) as a fixed effect. The average dwell time indeed was higher in EV than in both PH, b = 275.86, $CI_{95\%}$ = (258.45–293.27),[9] and MM, b = 342.77, $CI_{95\%}$ = (325.36–360.19); the average dwell time was also higher in PH than MM, b = 66.91, $CI_{95\%}$ = (49.50–84.33). We additionally checked whether the average dwell time is a valid measure to reflect differences of information acquisition and processing behavior during decision-making according to our strategies; it would be also plausible that a specific strategy leads to very long and very short dwell time, which might not change the average but the distribution. Therefore, we inspected histograms over dwell times revealing that dwell times were

[7]Schulte-Mecklenbeck data: EV-PH, b = 15.23, $CI_{95\%}$ = (9.21–21.25).

[8]As some readers might be also interested in response times: the same analysis on response times revealed similar results for the difference between EV-PH, b = 38.68, $CI_{95\%}$ = (36.49–40.87), EV-MM, b = 53.31, $CI_{95\%}$ = (51.12–55.51), and PH-MM, b = 14.63, $CI_{95\%}$ = (12.44–16.82), which is due to a high correlation between fixation frequencies and response times, r = 0.958, $CI_{95\%}$ = (0.954–0.961).

[9]Schulte-Mecklenbeck data: EV-PH, b = 3.02, $CI_{95\%}$ = (−3.18 to 9.23).

consistently longer in EV than in PH and MM as well as distributed similarly (see Figure 5).

In sum, results confirmed our predictions but contrast the finding of Schulte-Mecklenbeck, Kühberger, et al. (2017) concerning dwell times produced by PH and EV, which did not differ significantly.

4 | DISCUSSION

The investigation of cognitive processes requires process-tracing methodologies allowing for correct inferences of covert processes based on observable behavior. Almost all process-tracing studies in JDM apply a kind of reasoning called reverse inference (cf. Poldrack, 2006), and therefore rely on a close mapping between overt behavior and covert processes. As the mapping between overt and covert processes has rarely been evaluated, it is unclear how many recent process tracing studies committed the fallacy of reverse inference due to a violation of its prerequisite. The aim of our study is to scrutinize the results from Schulte-Mecklenbeck, Kühberger, et al. (2017) who identified problems with the prerequisite of reverse inference in risky choice due to mismatches of theoretically predicted and observed gaze behavior.

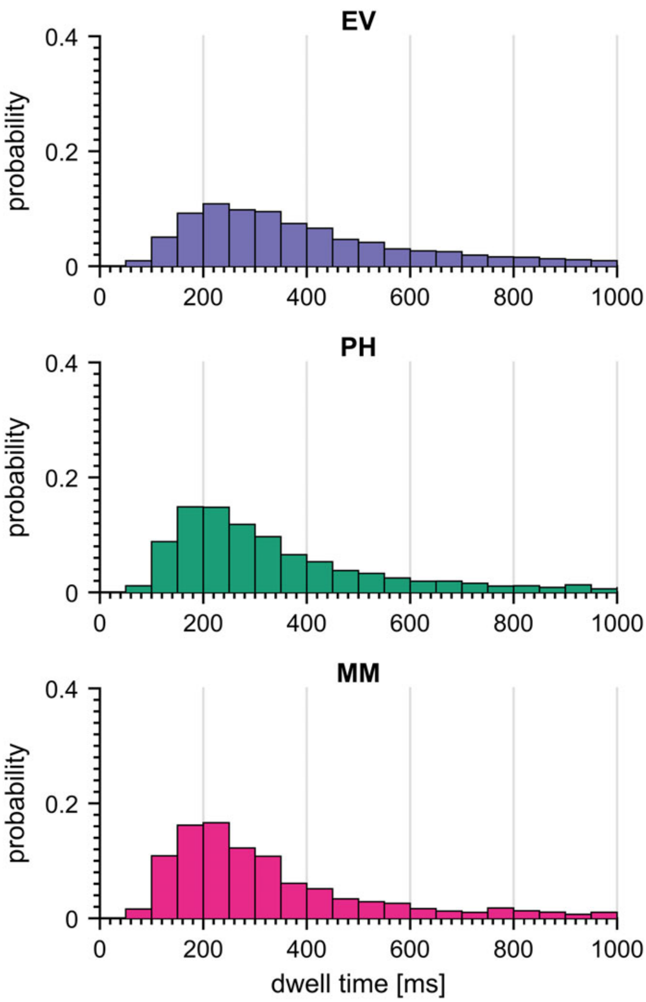

FIGURE 5 Histogram of the dwell time of all dwells for each of the three strategies. [Colour figure can be viewed at wileyonlinelibrary.com]

Therefore, we applied an updated forward inference paradigm that addresses three threats to the validity of the original paradigm used by Schulte-Mecklenbeck, Kühberger, et al. (2017): We used (a) gambles that allow different predictions for the used decision strategies in all cases; (b) a strategy check, which made sure participants executed the instructed decision strategy, and (c) rewarded correct execution of a given strategy. For the three instructed decision strategies, PH, EV, and MM, we investigated four dependent measures (choices, dwell patterns, dwell frequencies, and dwell times). In order to pursue our aim, those measures can be evaluated from two perspectives: First, the replication of the findings from Schulte-Mecklenbeck, Kühberger, et al. (2017), and second, the matching of theoretically predicted and empirically observed gaze behavior.

Concerning the replication of the previous finding, we found similar results for dwell patterns and overall dwell frequencies but not for dwell times; we found mixed results for the fine-grained analyses of dwell frequencies. Dwell patterns were attribute-wise when participants applied PH or MM; they were option-wise when participants applied EV. Dwell frequencies were highest when participants applied EV, followed by PH and MM, sequentially. Across strategies, outcomes were fixated more frequently than probabilities; large and small outcomes as well as large and small probabilities were fixated equally frequent. The former result is also consistent with previous work that also built on fine-grained predictions (Johnson et al., 2008) as well as work on the distribution of attention on outcomes and probabilities (Fiedler & Glöckner, 2012; Glöckner & Herbold, 2011; Pachur et al., 2013; Su et al., 2013), but contradicts findings on gaze behavior in proportion tasks that genuinely recruit a weighting and adding process, and hence predict equal distributions of fixations on outcomes and probabilities (Su et al., 2013). The latter, however, also deviates from the previous findings from Schulte-Mecklenbeck, Kühberger, et al. (2017), proposing more dwells on large outcomes and small probabilities. Dwell times were highest when participants applied EV, followed by PH and MM, sequentially, which also deviates from the previous findings proposing similar dwell times in EV and PH. Furthermore, we found longer dwells in general, which would associate gaze behavior in our study with a deeper level of processing, whereas gaze behavior in the previous study would be associated with a more superficial level of processing (Glöckner & Herbold, 2011; Velichkovsky, 1999). Despite those deviations from the previous study, we also found quantitatively more extreme results across the measures that were in line with the previous results. For dwell patterns, our results show an even stronger effect between EV and PH, which is driven by more attribute-wise transitions in PH; for fixation frequencies, the stronger effect between EV and PH is driven by less dwells in PH.

The observation of the more extreme effect in our study could be due to our improved forward inference paradigm consisting of a strategy check, a different mode of incentivizing, and a set of gamble problems in that each problem can be solved unambiguously applying either instructed strategy. By these means, our participants were urged and capacitated to apply the instructed decision strategies. Hence, we can likely rule out the possibility that participants eventually applied a simplified version of the instructed strategies, which does not apply to the Schulte-Mecklenbeck, Kühberger, et al. (2017) data. Indeed, comparing both data sets, the application of simplified versions of the strategy seems to be an adequate explanation for the quantitatively different effects when considering all dependent measures. In contrast, we have no evidence to believe that the varied presentation format might have introduced systematically different gaze behavior and hence caused the different effects (Smith & Krajbich, 2018).

Concerning the matching of theoretically predicted and empirically observed gaze behavior, we found that our results correspond to our predictions with respect to choices, dwell patterns, overall dwell frequencies, and dwell times, but differ partially (8 out of 30 predictions) with respect to the fine-grained analysis of dwell frequencies (see Table 2). As this perspective provides the targeted evaluation of the mapping between overt and covert processes, our results suggest that this prerequisite for reverse inference maintains for the most—but not for all—measures. From this point of view, reverse inference is in some occasions prone to produce fallacies due to a violation of its prerequisites. Specifically, if one chooses an inadequate measure on that the covert cognitive (decision) processes are reversely inferred.

With respect to EV, the prediction that the number of dwells on outcomes and probabilities should be equal did not match our results. Hence, we would suggest that an unequal number of dwell on outcomes and probabilities should not be taken as evidence against EV, and probably in favor of PH or MM. In fact, this mismatch has already been found in previous studies of risky choice using reverse-inference (Fiedler & Glöckner, 2012; Pachur et al., 2013; cf. Stewart, Hermens, & Matthews, 2015; Su et al., 2013). With respect to PH, the predictions that there should be more dwell on the minimum outcomes and probabilities than on the maximum outcomes and probabilities did also not match our results. Instead, such properties of the gamble problem were fixated equally often. Hence, we would suggest that an equal number of dwells on the maximum and minimum outcomes and probabilities should not be taken as evidence against PH. With respect to MM, we found that maximum and minimum outcomes were fixated equally often, which stands in a drastic contrast to the predictions. According to the predictions, the decision is solely based on the minimum outcomes, and hence, the maximum outcomes can be almost completely ignored. As this is not the case, we would argue that an equal number of dwell on minimum and maximum outcomes should be taken as evidence against MM.

4.1 | Limitations

Our conclusion, that reverse inference might be problematic in few occasions comes with some limitations we divide into methodological and theoretical ones.

4.1.1 | Methodological limitations

The methodological limitations target the artificial setting of a forward inference task as well as our set of instructed decision strategies. In our forward inference paradigm, we teach people the algorithmic representation of decision strategies and we must rely on a full internalization of this representation. Certainly, we cannot rule out the possibility that some other processes are going on that are unrelated to the decision process such as retrieval processes for the instructed strategy. However, we must also rely on the same assumption when people spontaneously think about gambles. In our paradigm, we urge people to process the algorithmic representation of the decision strategy very closely, whereas in uninstructed settings, people can process what and when they want. This brings us to the claim that such an artificial setting might not the perfect, but a better way to examine the mapping between covert and overt processes.

For our paradigm, we chose three different decision strategies (maximizing the EV, the PH, and the MM). As for the former two, we chose the same strategies as in the Schulte-Mecklenbeck, Kühberger, et al. (2017) study, because those strategies are well-known among decision scientists, make fundamentally different process predictions, and are trainable. However, we suspected that people might deviate from processing the algorithmic representation due to many calculation operations that are involved. To overcome this issue, we additionally introduced the MM as the simplest possible decision strategy for that we did not see any a priori reason that people might deviate from processing the algorithmic representation closely. Hence, given the mere amount of decision strategies in the literature, our three strategies pose one possible subsample that pictures the variety of strategies with regard to their difficulty (e.g., absolute amount of required EIPs) and complexity (e.g., amount of different involved EIPs) and are yet teachable. By choosing those strategies, we make no claim about the plausibility of those strategies describing the true decision process. Our reasoning was purely motivated to examine the mapping between covert and overt processes in a forward inference paradigm.

For our paradigm, we also chose gambles that could be solved applying either of the three different strategies. This constraint consequently limits the selection of gambles to a small range that might influence the generalizability of our results. We have applied this constraint in order to provide each of the three decision strategies with a clear termination and to prevent any subsequent decision process and hence gaze behavior that was not instructed. Thus, this feature of our paradigm should have increased the mapping between covert and overt processes. In terms of generalizability, we would expect an even less close mapping than we have observed when applying a more representative gamble set.

4.1.2 | Theoretical limitations

We evaluate the mapping between overt and covert processes by comparing theoretically predicted and empirically observed gaze behavior. By doing so, our evaluation and hence our conclusion crucially depends on the quality of the employed predictions. Therefore, the hitherto existing interpretation of our and the previous results must be contrasted with the alternative explanation that we just used poor predictions for the gaze behavior that should be observed given our decision processes. The mere fact that we only used canonical predictions that have already been employed in the recent literature cannot invalidate this limitation completely. Though, questioning the quality of the predictions we commonly use in eye-tracking studies, sketches a promising future application of our forward inference paradigm.

4.2 | Future directions

Assuming that reverse inference is unproblematic given good predictions, forward inference can be applied to check and improve the quality of the hitherto used prediction. From this perspective, our results suggest that some of the canonical predictions concerning dwell frequencies are poor, though they yield some validity with respect to the differentiation between decision processes. This becomes evident by employing a data-driven machine learning approach to learn from the forward-inference data the mapping between covert cognitive processing and overt gaze behavior (for related approach, please see Król & Król, 2017). Overall, the results show that our decision processes can be classified with high accuracy using different combinations of gaze measure. Furthermore, the results show that the dwell distribution on specific gamble properties as used in the fine-grained analysis of dwell frequencies increases the classification accuracy when they are used exclusively or in combination with two of the other measures. However, the dwell patterns and the dwell frequencies seem to produce the strongest combination of gaze measures to discriminate our decision strategies (for more details, see Appendix B). Hence, different decision processes lead to different gaze behaviors, which is revealed across all gaze measures. In order to rehabilitate reverse inference, forward inference is a promising tool to identify poor predictions and to convert them into accurate predictions.

By this application, forward inference could help to convey JDM research into a more proces-orientated paradigm (Oppenheimer & Kelso, 2015; Scherbaum, Dshemuchadse, & Kalis, 2008; Schulte-Mecklenbeck, Johnson, et al., 2017). In Marr's (1982) terminology, forward inference could help to switch decision theory from the computational level, where problems are specified in the generic manner to the algorithmic level describing how exactly the computational problems are solved. Hence, forward inference provides a framework in order to obtain that switch as it delivers process data from specific cognitive processes that can in turn be used to build models of decision-making on the algorithmic level or to refine computational models towards the algorithmic level.

5 | CONCLUSION

Our study demonstrates that a forward inference paradigm is a useful framework to evaluate the mapping between overt information processing as indicated by gaze behavior and covert cognitive processes as predicted by decision-making models. Our experiment was designed to scrutinize the Schulte-Mecklenbeck, Kühberger, et al. (2017) findings applying an improved forward inference paradigm. Overall, we found mixed results with respect to the validation of the previous findings, though we draw a similar but more accentuated conclusion: our findings shed doubts on recent JDM research having committed the fallacy of the reverse inference due to a violation of its prerequisite. Even though the mapping between gaze decision processes seems sufficiently high, the identification of decision processes crucially depends on the quality of the model's predictions. We identified several predictions that might lead to erroneous reverse inferences. Therefore, the usage of stronger predictions would restore the prerequisite and could hence rehabilitate reverse inference. To accomplish this aim, our improved forward inference paradigm might be an adequate methodology.

6 | CONTEXT OF THE RESEARCH

The research reported in this article originated when the first, third, and last author searched for eye-movement data in which it was clear what process has been applied during decision-making. The second author yielded such a data set and shared his data and manuscript (at that time submitted). The first author identified the provided data set to be unsuitable for investigating his initial research question due to methodological limitations as discussed in this article. The first, third, and last author decided to collect more suitable data by applying an improved version of the second author's paradigm. The first author planned and programmed the new paradigm based on the initially shared manuscript. The first author also managed data collection and conducted data analysis. By then, the second author's manuscript had been accepted for publication (see Schulte-Mecklenbeck, Kühberger, et al., 2017). As the second author is an expert in the field and since was involved in the new work by his early sharing of unpublished data and manuscript, the first and last authors invited the second author to support the current work as a co-author. Final editing of the manuscript was done by all authors.

ACKNOWLEDGEMENTS

This research was funded by the Department of Psychology, Technische Universität Dresden (startup grant awarded to MS). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Andreas Glöckner, Ulrike Senftleben, and an anonymous reviewer for providing helpful comments on an earlier version of this manuscript and Marie Gotthardt and Judith Herbers for their support in data collection.

ORCID

Martin Schoemann https://orcid.org/0000-0002-2531-4175
Michael Schulte-Mecklenbeck https://orcid.org/0000-0002-0406-8809
Frank Renkewitz https://orcid.org/0000-0001-8072-6802
Stefan Scherbaum https://orcid.org/0000-0002-4408-6016

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought*. New York: Psychology Press.

Bettman, J. R., Johnson, E. J., & Payne, J. W. (1990). A componential analysis of cognitive effort in choice. *Organizational Behavior and Human Decision Processes*, *45*, 111–139. https://doi.org/10.1016/0749-5978(90)90007-V

Böckenholt, U., & Hynan, L. S. (1994). Caveats on a process-tracing measure and a remedy. *Journal of Behavioral Decision Making*, *7*(2), 103–117. https://doi.org/10.1002/bdm.3960070203

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409–432. https://doi.org/10.1093/acprof:oso/9780199744282.003.0007

Brandstätter, E., & Körner, C. (2014). Attention in risky choice. *Acta Psychologica*, *152*, 166–176. https://doi.org/10.1016/j.actpsy.2014.08.008

Bröder, A. (2010). Outcome-based strategy classification. In A. Glöckner, & C. L. Witteman (Eds.), *Foundations for tracing intuition: Challenges and methods* (pp. 61–82). London: Psychology Press & Routledge.

Coombs, C. H., Dawes, R. M. D., & Tversky, A. (1970). *Mathematical Psychology*. New Jersey: Prentice-Hall.

Fiedler, S., & Glöckner, A. (2012). The dynamics of decision making in risky choice: An eye-tracking analysis. *Frontiers in Psychology*, *3*, 335. https://doi.org/10.3389/fpsyg.2012.00335

Forstmann, B. U., Wagenmakers, E.-J., Eichele, T., Brown, S. D., & Serences, J. T. (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: Opposites attract? *Trends in Cognitive Sciences*, *15*(6), 272–279. https://doi.org/10.1016/j.tics.2011.04.002

Glaholt, M. G., & Reingold, E. M. (2011). Eye movement monitoring as a process tracing methodology in decision making research. *Journal of Neuroscience, Psychology, and Economics*, *4*(2), 125–146. https://doi.org/10.1037/a0020692

Glöckner, A., & Herbold, A.-K. (2011). An eye-tracking study on information processing in risky decisions: Evidence for compensatory strategies based on automatic processes. *Journal of Behavioral Decision Making*, *24*(1), 71–98. https://doi.org/10.1002/bdm.684

Glymour, C., & Hanson, C. (2016). Reverse inference in neuropsychology. *British Journal for the Philosophy of Science*, *67*(4), 1139–1153. https://doi.org/10.1093/bjps/axv019

Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns. *Vision Research*, *62*, 1–8. https://doi.org/10.1016/j.visres.2012.03.019

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 - A guide for the organization of experiments in economics. *Working Paper Series in Economics*, 1–67. Retrieved from http://ideas.repec.org/p/kls/series/0010.html

Hacking, I. (1984). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge: Cambridge University Press.

Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, *9*(4), 188–194. https://doi.org/10.1016/j.tics.2005.02.009

Heit, E. (2015). Brain imaging, forward inference, and theories of reasoning. *Frontiers in Human Neuroscience*, *8*, 1056. https://doi.org/10.3389/fnhum.2014.01056

Henson, R. (2006). Forward inference using functional neuroimaging: Dissociations versus associations. *Trends in Cognitive Sciences*, *10*(2), 64–69. https://doi.org/10.1016/j.tics.2005.12.005

Johnson, E. J., & Ratcliff, R. (2014). Computational and process models of descision making in psychology and behavioral economics. In P. Glimcher, & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain* (2, pp. 35–47). Amsterdam: Acad. Press.

Johnson, E. J., Schulte-Mecklenbeck, M., & Willemsen, M. C. (2008). Process models deserve process data: Comment on Brandstätter, Gigerenzer, and Hertwig (2006). *Psychological Review*, *115*(1), 263–272. https://doi.org/10.1037/0033-295X.115.1.263

Król, M. E., & Król, M. (2017). A novel approach to studying strategic decisions with eye-tracking and machine learning. *Judgment and Decision making*, *12*(6), 596–609. https://search.proquest.com/openview/11620a0314a50879830e04f1f5f024bf/1?pq-origsite

Machery, E. (2014). In defense of reverse inference. *The British Journal for the Philosophy of Science*, *65*(2), 251–267. https://doi.org/10.1093/bjps/axs044

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Oppenheimer, D. M., & Kelso, E. (2015). Information processing as a paradigm for decision making. *Annual Review of Psychology*, *66*(1), 277–294. https://doi.org/10.1146/annurev-psych-010814-015148

Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. https://doi.org/10.1016/j.actpsy.2013.06.003

Pachur, T., Hertwig, R., Gigerenzer, G., & Brandstätter, E. (2013). Testing process predictions of models of risky choice: A quantitative model comparison approach. *Frontiers in Psychology*, *4*(September), 1–22. https://doi.org/10.3389/fpsyg.2013.00646

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(3), 534–552.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Poldrack, R. A. (2006). Can cognitive processes be inferred from neuroimaging data? *Trends in Cognitive Sciences*, *10*(2), 59–63. https://doi.org/10.1016/j.tics.2005.12.004

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge: MIT Press.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria. https://www.r-project.org/

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Scherbaum, S., Dshemuchadse, M., & Kalis, A. (2008). Making decisions with a continuous mind. *Cognitive, Affective, & Behavioral Neuroscience*, *8*(4), 454–474. https://doi.org/10.3758/CABN.8.4.454

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-tracing methods in decision making: On growing up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450. https://doi.org/10.1177/0963721417708229

Schulte-Mecklenbeck, M., Kühberger, A., Gagl, B., & Hutzler, F. (2017). Inducing cognitive processes: Bringing process measures and cognitive processes closer together. *Journal of Behavioral Decision Making*, *30*(5), 1001–1013. https://doi.org/10.1002/bdm.2007

Smith, S. M., & Krajbich, I. (2018). Attention and choice across domains. *Journal of Experimental Psychology. General*, *147*, 1810–1826. https://doi.org/10.1037/xge0000482

Stewart, N., Hermens, F., & Matthews, W. J. (2015). Eye movements in risky choice. *Journal of Behavioral Decision Making*, *29*(2-3), 116–136. https://doi.org/10.1002/bdm.1854

Su, Y., Rao, L.-L., Sun, H.-Y., Du, X.-L., Li, X., & Li, S. (2013). Is making a risky choice based on a weighting and adding process? An eye-tracking investigation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(6), 1765–1780. https://doi.org/10.1037/a0032861

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & Van Maanen, L. (2017). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology*, *76*, 65–79. https://doi.org/10.1016/j.jmp.2016.01.001

Turner, B. M., Rodriguez, C. A., Norcia, T. M., McClure, S. M., & Steyvers, M. (2016). Why more is better: Simultaneous modeling of EEG, fMRI, and behavioral data. *NeuroImage*, *128*, 96–115. https://doi.org/10.1016/j.neuroimage.2015.12.030

Velichkovsky, B. M. (1999). From levels of processing to stratification of cognition: Converging evidence from three domains of research. In B. H. Challis, & B. M. Velichkovsky (Eds.), *Stratification in cognition and consciousness* (pp. 203–235). Amsterdam: John Benjamins Publishing Company.

Zhou, L., Zhang, Y.-Y., Wang, Z.-J., Rao, L.-L., Wang, W., Li, S., … Liang, Z.-Y. (2016). A scanpath analysis of the risky decision-making process. *Journal of Behavioral Decision Making*, *29*(2–3), 169–182. https://doi.org/10.1002/bdm.1943

AUTHOR BIOGRAPHIES

Martin Schoemann is a PhD student at the Technische Universität Dresden and associated with the Collaborative Research Center 940 “Volition and Cognitive Control”. His research focuses on the cognitive processes underlying decision-making and process-tracing methods such as eye tracking and mouse cursor tracking.

Michael Schulte-Mecklenbeck is a senior lecturer at the University of Bern and an associated Research Scientist at the Max Planck Institute for Human Development, Berlin. His research focuses on process tracing, food choice, heuristics, and consumer behavior.

Frank Renkewitz is a lecturer at the University of Erfurt, Germany. His research focuses on decision-making and medical decision-making, process-tracing, and research methods.

Stefan Scherbaum is Professor of Research Methods and Cognitive Modelling at the Technische Universität Dresden, Germany. His research focuses on process tracing, decision-making, cognitive control, and social decision-making.

How to cite this article: Schoemann M, Schulte-Mecklenbeck M, Renkewitz F, Scherbaum S. Forward inference in risky choice: Mapping gaze and decision processes. *J Behav Dec Making*. 2019;1–15. https://doi.org/10.1002/bdm.2129

APPENDIX A

TABLE A1 The 21 gamble problems presented in the paradigm

	Gamble A				Gamble B					
	Option 1		Option 2		Option 1		Option 2			
Number	O1	P1	O2	P1	O1	P1	O2	P1	EV-R	Reasons
1	5	60	8	40	2	80	10	20	1,72	1
2	3	70	8	30	4	60	9	40	0,75	1
3	4	60	5	40	3	60	4	40	1,29	1
4	2	90	20	10	5	60	10	40	0,54	1
5	5	80	20	20	2	80	10	20	2,22	1
6	12	60	18	40	10	60	15	40	1,20	1
7	4	60	10	40	2	80	9	20	1,88	1
8	5	70	11	30	4	80	9	20	1,36	2
9	10	60	20	40	9	70	18	30	1,20	2
10	9	80	20	20	10	70	18	30	0,90	2
11	3	70	14	30	2	60	16	40	0,83	2
12	9	60	18	40	8	70	15	30	1,25	2
13	9	80	20	20	10	60	18	40	0,85	2
14	9	80	20	20	10	60	18	40	0,85	2
15	5	75	11	25	4	80	8	20	1,35	3
16	10	60	20	40	9	65	18	35	1,15	3
17	5	80	19	20	6	75	11	25	1,08	3
18	3	70	13	30	2	65	15	35	0,92	3
19	9	75	18	25	8	70	15	30	1,11	3
20	8	80	20	20	9	75	18	25	0,92	3
21	9	80	20	20	10	75	17	25	0,95	3

Note. Outcomes (O) are given in Euro (€) and Probabilities (P) are given in percent (%). EV-R denominates the ratio of the expected values of both gamble options.

APPENDIX B

In the introductory part of this article, we argued that a forward inference paradigm can be used to evaluate the mapping between covert cognitive processing and overt observable behavior. Beside the mere evaluation of the mapping, forward inference paradigm can also be applied to adopt a more data-driven approach to enhance our reverse inference on cognitive processes. In this respect, a machine learning approach might be a promising path in order to learn from the forward-inference data what eye-movement measures best characterize the different cognitive processes. Eventually, the trained classifier could be used to identify cognitive processes in reverse-inference data (cf. Greene, Liu, & Wolfe, 2012).

As a first step in this direction, we trained linear support vector machines fed with various combinations of standardized eye-movement measures using a 10 fold cross-validation (see Table B1 and Figure B1). In this validation method, the data are partitioned into a specific number of bins (i.e., folds). The classifier is trained for each fold using all the data outside the respective fold; the performance of the classifier is in turn tested using the data inside the fold. The predictive accuracy of the final classifier is the average accuracy over all folds. The generation of the dataset as well as the fitting and validation of the classifiers is covered in our analysis script (available online).

TABLE B1 Overview of the predictive accuracy and the TPR of trained linear SVMs fed with various combinations of standardized eye-movement measures

Dwell pattern	Dwell frequency	Dwell time	Fine-grained analyses*	Classification accuracy (in %)	TPR (in %)		
					EV	PH	MM
X	–	–	–	70.10	98.77	80.65	30.88
–	X	–	–	71.12	45.31	73.89	94.16
–	–	X	–	51.41	70.51	15.67	68.05
–	–	–	X	**82.23**	95.39	62.21	89.09
X	X	–	–	**93.65**	99.39	88.33	93.24
X	–	X	–	79.77	98.46	71.27	69.59
X	–	–	X	91.91	98.92	87.25	89.55
–	X	X	–	83.56	83.72	73.58	93.39
–	X	–	X	86.23	94.01	70.66	94.01
–	–	X	X	83.46	90.94	70.20	89.25
X	X	X	–	94.06	99.08	89.71	93.39
X	X	–	X	**94.67**	99.08	90.63	94.32
X	–	X	X	92.17	98.92	88.02	89.55
X	X	X	X	**-- 95.03 --**	99.08	91.40	94.62

Note. The predictor "fine-grained analyses" summarizes the inclusion of four additional predictors: the standardized ratio of dwells on O_{max}, O_{min}, P_{max}, and P_{min}.

Abbreviations: EV, expected value; MM, minimax heuristic; PH, priority heuristic; SVM, support vector machine; TPR, true positive rate.

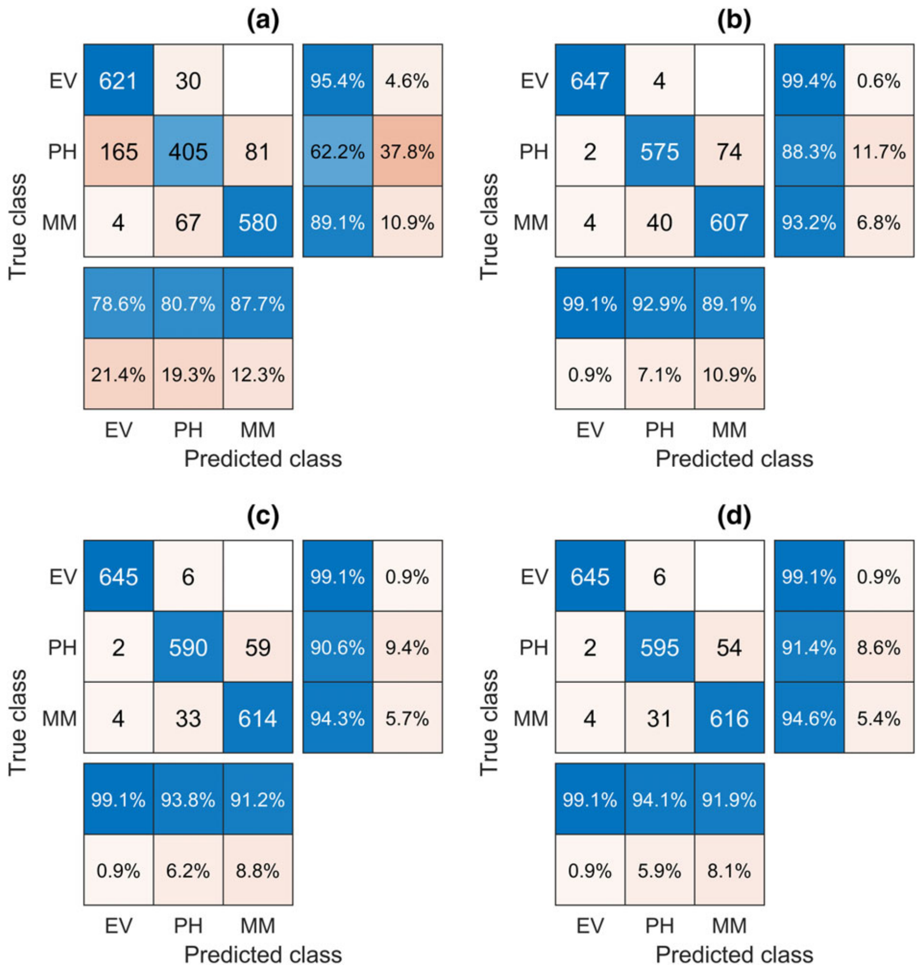

FIGURE B1 Confusion matrices for the best performing classifier within each number of predictor combinations as printed in bold in Table B1. Panels A–D matches the order from top to bottom in Table B1. The true positive rate and the false negative rate are depicted on the right side of each panel; the positive predictive values and the false discovery rates are depicted below each panel. Note: the confusion matrices for the remaining classifiers can easily be plotted using our analysis script available online. [Colour figure can be viewed at wileyonlinelibrary.com]