Journal of Behavioral Decision Making, *J. Behav. Dec. Making* (2015)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/bdm.1925

The (In)visibility of Psychodiagnosticians' Expertise

MICHAEL SCHULTE-MECKLENBECK[1,2]*, NANON L. SPAANJAARS[3] and CILIA L. M. WITTEMAN[3]

[1]*Department of Business Administration, University of Bern, Switzerland*
[2]*Max Planck Institute for Human Development, Berlin, Germany*
[3]*Behavioural Science Institute, Radboud University, Nijmegen, Netherlands*

ABSTRACT

This study investigates decision making in mental health care. Specifically, it compares the diagnostic decision outcomes (i.e., the quality of diagnoses) and the diagnostic decision process (i.e., pre-decisional information acquisition patterns) of novice and experienced clinical psychologists. Participants' eye movements were recorded while they completed diagnostic tasks, classifying mental disorders. In line with previous research, our findings indicate that diagnosticians' performance is not related to their clinical experience. Eye-tracking data provide corroborative evidence for this result from the process perspective: experience does not predict changes in cue inspection patterns. For future research into expertise in this domain, it is advisable to track individual differences between clinicians rather than study differences on the group level. Copyright © 2015 John Wiley & Sons, Ltd.

Additional supporting information may be found in the online version of this article at the publisher's web-site.

KEY WORDS eye-tracking; diagnostic decision making; experience; clinical psychology

The role of expertise in diagnostic decision making has been studied for nearly four decades. One set of results provides evidence that more experienced clinicians are more competent in applying categorization rules and, thus, making better classifications than novices (Brammer, 2002; Kim & Ahn, 2002). Another set of studies shows evidence that there is no substantial difference in the accuracy of diagnostic decisions (e.g. Ægisdóttir *et al*., 2006; Garb, 1998; Strasser & Gruber, 2004; Witteman & Van den Bercken, 2007). In these studies, experienced clinicians typically showed relatively low levels of accuracy in diagnosing mental disorders (Brailey, Vasterling, & Franks, 2001), yet they were overconfident about the accuracy of their choices (Croskerry & Norman, 2008; Hogarth, 2010; Menkhoff, Schmeling, & Schmidt, 2013). Spengler *et al*. (2007) performed a meta-analysis of published work on diagnostic accuracy and concluded that there is a reliable but small effect ($d = 0.12$) in favour of experienced over less experienced clinicians.

Of course, the accuracy of a decision is only one side of the issue. Even if experienced clinicians' accuracy is only slightly higher than that of novices, it seems reasonable to assume that how this group decides—that is, how they process information—is influenced by their experience. It has indeed been shown that decision-making processes change with experience (Elstein & Schwartz, 2002). Medical clinicians with more experience seem to use more encapsulated knowledge and to make their decisions faster than novices do (Schmidt & Rikers, 2007). Importantly, little is known about how mental health clinicians' cognitive processes change with experience, except that their memories become less detailed and more abstract than those of novices (Brailey *et al*., 2001; Witteman & Tollenaar, 2012).

*Correspondence to: Michael Schulte-Mecklenbeck, Department of Business Administration, University of Bern, Engehaldenstr. 4, 3012 Bern, Switzerland. E-mail: research@schulte-mecklenbeck.com

To our knowledge, the present study is the first to use eye-tracking methods to evaluate clinical psychologists' decision making processes (for an overview of process tracing methods see Schulte-Mecklenbeck, Kühberger, & Ranyard, 2011a, 2011b). Eye-tracking might thereby provide a unique method for evaluating differences between novice and experienced participants. First, process tracing methods that rely on verbal utterances, for example, thinking aloud (Ranyard & Svenson, 2011; Russo, Johnson, & Stephens, 1989), might give less objective results, because more experience leads to more encapsulated knowledge (Schmidt & Rikers, 2007), which is harder to verbalise. This would favour novice clinicians, who would find it easier to say what they think, over experienced clinicians. Second, a meta-analysis of eye-tracking studies in the medical domain shows that this may be a promising avenue to find experience-related differences: more experienced doctors were found to have shorter fixation durations, more fixations on task-relevant areas and fewer fixations on task-redundant areas than novices (Gegenfurtner, Lehtinen, & Säljö, 2011).

We investigate experience-related differences between novice and experienced clinical psychologists by comparing not only the accuracy of their diagnostic decisions (i.e. decision outcomes) but also their processes of information acquisition (i.e. decision processes), using eye-tracking.

Hypotheses

For decision outcomes, we hypothesized—based on the meta-analytic findings of Spengler *et al*. (2007)—that experienced clinical psychologists would be more accurate than novices in diagnosing cases. Furthermore, we expected that experienced clinicians would overestimate their number of correct responses more often than novices would (Croskerry & Norman, 2008).

For decision processes, we hypothesized that experienced clinical psychologists, like their colleagues in other medical domains, would make decisions faster than novices (Schmidt & Rikers, 2007). Likewise, in line with findings in the medical professions, we expected experienced mental-health clinicians to have longer dwell times on task-relevant (diagnostic) information and shorter dwell times on task-redundant (nondiagnostic) information, relative to novices (Gegenfurtner *et al.*, 2011).

METHOD

Participants

Participants (50 participants: 29 novice; 2 men; 24.6 years; $SD=2.3$ years and 21 experienced; 6 men; 38.1 years; $SD=9.6$ years; see the Supporting Information for details) were recruited through a convenience sample and received €10 as a showup fee. The novices were master's students with no professional experience apart from a clinical internship included in their degree course. The experienced clinicians had an average of 11.1 years ($SD=7.8$ years; $Mdn=11$ years; range: 2–26 years) experience in the profession; they worked either in private practice or in an institute for mental healthcare in the Netherlands.

Materials

In each trial, participants were presented with eight symptoms and asked to choose which of two mental disorders they best matched. Of the eight symptoms presented in each trial, the number of diagnostic criteria is ranged from one to four (see Supporting Information for details of how the stimulus material was selected and the eye-tracking data were recorded).

Procedure

Clinicians first completed a computerized questionnaire (Inquisit, 2009) assessing their experience in the profession. All participants then received instructions about the task and the stimulus setup and were calibrated with the eye tracker. In the main diagnostic task, participants were asked to indicate, as soon as they were certain of their decision, which of two diagnostic labels (see Supporting Information for details) best fit the§§ case information by pressing a key on either the left ("A") or the right side ("L") of the keyboard. Participants were instructed to do this task as quickly and accurately as possible, but no time limit was imposed. When participants finished the 22 trials, the eye tracker was removed, and they were asked how many of their classifications they thought were correct (Gigerenzer, Hoffrage, & Kleinbölting, 1991). Each trial was then shown again, in a different fixed randomized order, and participants were again asked to decide which diagnosis was applicable, but without their eye movements being tracked. On average, the experiment was completed in 45 to 60 min.

RESULTS

Outcome analysis

Accuracy

We evaluated accuracy (number of correctly identified diagnoses) at T1 and T2. On average across the 22 trials, the experienced clinicians made 66% correct decisions at T1 and 65% correct decisions at T2. Averaging the z-transformed correlations per person between T1 and T2 resulted in $z'=0.46$ for the experienced group. A similar picture emerged in the novice clinician group, with 63% correct decisions at T1 and 65% correct decisions at T2. On average, novices were more consistent than experienced clinicians, with $z'=0.58$. Comparison of the z-scores revealed that there was no significant difference between novices and experienced clinicians in their accuracy ($z=-0.58$, $p=0.56$).

We estimated a multilevel logistic regression[1] predicting correct response with "participants" and "tasks" as random intercepts, and "expertise" (experienced versus novice) and "number of diagnostic cues" (one, two, three or four) as fixed effects. The interaction term for "expertise" and "number of diagnostic cues" was also included as a fixed effect. Neither expertise, $b=0.29$, 95% confidence interval (CI) [−0.35, 0.93], nor number of diagnostic cues, $b=0.07$, 95% CI [−0.46, 0.61], resulted in a significant effect. The interaction between "expertise" and "number of diagnostic cues" also showed no significant effect, $b=-0.24$, 95% CI [−0.56, 0.06]. We conclude that neither experience nor number of diagnostic cues is predictive of number of correct responses in the classification task.[2]

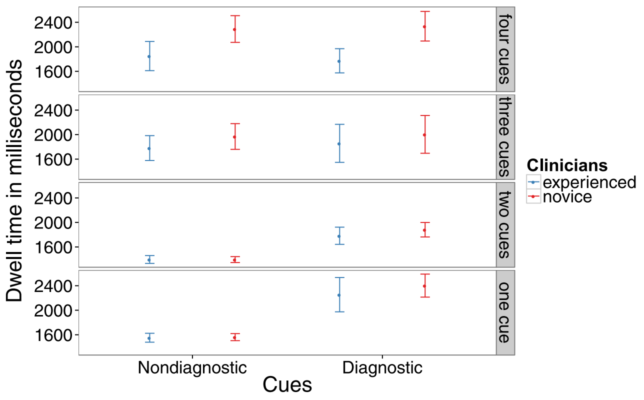

Figure 1. Average dwell time for diagnostic and nondiagnostic cues in the one, two, three, and four diagnostic cues conditions. Data are presented separately for experienced and novice clinicians. Error bars show 95% confidence intervals

Process analysis

On average, participants fixated 29 times, $SD=16$ times, on the eight cues and two disorders (see Figure S1 in the Supporting Information), which took them on average 15 s, $SD=9$ s, per trial.

As Figure 1 shows, when fewer diagnostic cues were available (one or two), more time was spent on diagnostic than on nondiagnostic cues. When the number of diagnostic cues was higher (three or four), this difference diminished. Importantly, we found no difference in dwell time as a main effect or as an

[1] All of the following analyses were conducted using a mixed-model approach with the lmer and glmer functions of the lme4 package (Bates, Maechler, Bolker, & Walker, 2014).

[2] See the Supporting Information for details of a calibration analysis of experienced versus novice clinicians, indicating that there is no difference between over- and underconfidence in choices between groups.

DOI: 10.1002/bdm

interaction with the factors tested above (i.e. expertise and number of diagnostic cues).

We estimated a multilevel regression predicting dwell time with "participants" and "tasks" as random intercepts and "expertise" (experienced versus novice), "diagnosticity" (diagnostic versus nondiagnostic) and "number of diagnostic cues" (one, two, three and four) as fixed effects. The interaction term for "expertise", "diagnosticity" and "number of diagnostic cues" was also included as a fixed effect.[3] We found no effect of expertise on total dwell time, $b = 0.01$, 95% CI [−0.17, 0.20], or number of diagnostic cues, $b = 0.08$, 95% CI [−0.05, 0.22]. Diagnostic cues were looked at significantly longer than nondiagnostic cues (diagnostic: $M = 2012$ ms, $SD = 1554$; nondiagnostic: $M = 1530$ ms, $SD = 1220$ ms, $b = 0.61$, 95% CI [0.48, 0.72]). A significant interaction was found between diagnosticity of cues and number of diagnostic cues, $b = -0.18$, 95% CI [−0.23, −0.12], indicating that the difference on dwell time between diagnostic and nondiagnostic cues is lower with fewer cues.

Next, we investigated whether task completion times differed for correct and incorrect responses and how these patterns changed over time. We estimated a multilevel regression predicting task completion time with "participants" and "tasks" as random intercepts and "expertise" (experienced versus novice), "correctness" (correct versus incorrect response) and "task position" (from 1 to 22) as fixed effects. The interaction term for "expertise", "correctness" and "task position" was also included as a fixed effect. Task position was significant, indicating that the average time spent on a task decreased significantly across the 22 trials, $b = -0.29$, 95% CI [−0.50, −0.09], from an initial 15.2 s, $SD = 1.1$ s, in task 1 to 7.8 s, $SD = 4.3$ s, in task 22 (Figure 2). Neither expertise, $b = 0.88$, 95% CI [−1.63, 3.39], nor number of correct response, $b = -0.56$, 95% CI [−1.72, 0.60], or any other of the interactions resulted in a significant effect.

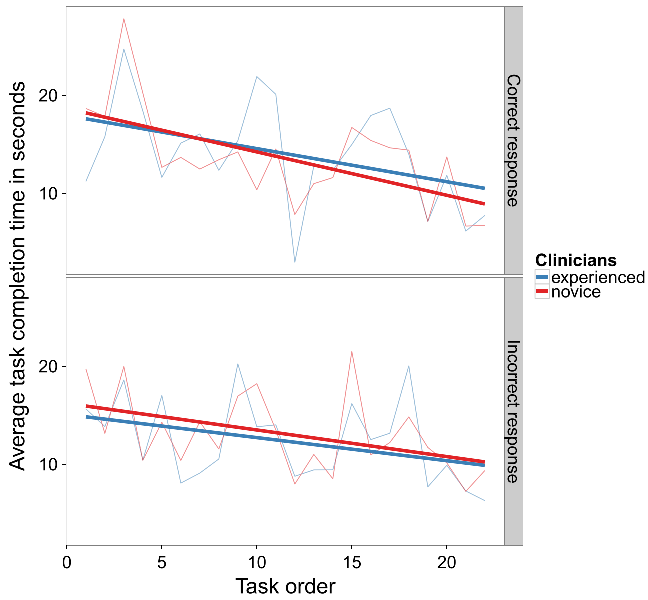

Figure 2. Task completion time for correct and incorrect responses across the 22 tasks. Data are presented separately for experienced and novice clinicians across the 22 tasks

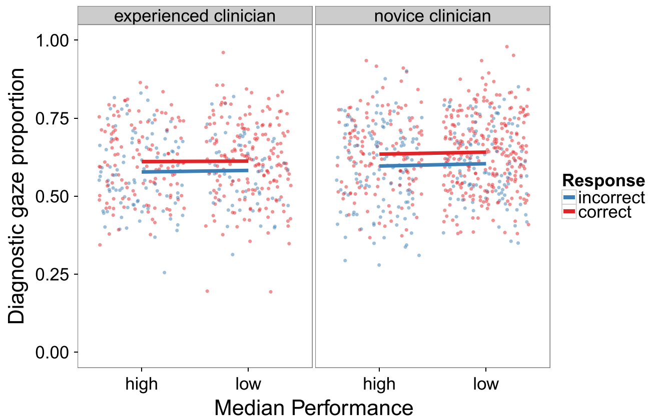

Figure 3. Diagnostic gaze proportion for incorrect and correct responses. The display shows experienced and novice clinicians. Each group is split on number of correct responses into a high performance (above median) and low performance (low median) group

Choices and dwell time on diagnostic information

To evaluate whether there is a difference between experienced and novice clinicians in the attention to diagnostic versus nondiagnostic items, we calculated the diagnostic gaze proportion.[4] This measure follows the biased sampling analysis reported in Ashby, Dickert, and Glöckner (2012). It is the proportion of dwell time spent on diagnostic cues relative to the overall dwell time on diagnostic and nondiagnostic cues.

As Figure 3 shows, the average diagnostic gaze proportions for the two groups, experienced clinicians and novice clinicians, were similar, regardless of whether responses were correct or incorrect or overall performance high or low. The centres of the two distributions are shifted upwards, above the 0.50 ratio, indicating overall more attention to diagnostic than nondiagnostic cues, regardless of response quality (correct versus incorrect).

To qualify this observation, we estimated a multilevel regression predicting diagnostic gaze proportion with "participants" and "tasks" as random intercepts and "expertise" (experienced versus novice), "correctness" (correct versus incorrect response) and "performance" (high versus low[5]) as fixed effects. The interaction term for "expertise", "correctness" and "performance" was also included as a fixed effect. We found a significant main effect of "correctness", $b = 0.04$, 95% CI [0.01, 0.06], indicating that, for correct responses, a higher diagnostic gaze proportion resulted. Neither "expertise", $b = 0.01$, 95% CI [−0.001, 0.04], nor "performance", $b = 0.008$, 95% CI [−0.02, 0.04], or any of the interactions

[3]To account for skewness of the dataset, we use log-transformed dwell times as a dependent measure in all the analyses reported here. All tests were also run with the number of acquisitions (i.e. fixations) as a dependent measure, rendering the same results.

[4]The diagnostic gaze proportion is calculated as follows: first, we normalized gaze duration on diagnostic and nondiagnostic cues by the number of cues in each task giving us four difficulty levels (1, 2, 3 and 4 diagnostic and 7, 6, 5 and 4 nondiagnostic, respectively). For this normalized gaze duration, we then calculated the proportion of dwell time spent on diagnostic cues relative to the overall dwell time on diagnostic and nondiagnostic cues for each participant, for each task.

[5]Performance was calculated based on a median split on correct responses separately for experienced and novice clinicians.

DOI: 10.1002/bdm

resulted in a significant effect. This indicates that none of the predictors, except correctness of the response, explained the lack of difference in gaze proportion between the two groups.

DISCUSSION

Tracey, Wampold, Lichtenberg, and Goodyear (2014) recently raised the question whether expertise in psychotherapy is an "elusive goal". This question was motivated by the observations of many decades of research on the failure of more experienced groups to outperform novices. We add a new perspective to this question by using eye-tracking data to investigate the diagnostic decision processes of clinical psychologists with different levels of experience. Our outcome hypothesis—that clinicians' decisions would become more accurate with experience (Spengler *et al*., 2007)—was not supported by our data. To our surprise, consistency between the diagnostic decisions made during eye-tracking and shortly afterwards, in a second round of testing, was low in both groups. Given such low correlations, participants must have switched both from correct to incorrect diagnoses and vice versa. Calculations of diagnostic accuracy on a group level may therefore be misleading (see Limitations in the Supporting Information).

We had also expected calibration to be poorer for experienced clinicians, who we predicted to be overconfident in their choices (Einhorn & Hogarth, 1978; Friedlander & Phillips, 1984; Garb, 1986). In fact, both groups showed the same level of overconfidence (see Supporting Information for details). Experienced clinicians did not become any more confident in their diagnoses, which points to some awareness of their diagnostic abilities (or lack thereof).

With respect to the process data, we had expected experienced clinical psychologists to be faster in their judgements and more focused on task-relevant, diagnostic cues. However, we found that experienced clinicians did not differ from novices in how they acquired information. Both groups looked at diagnostic items longer than at nondiagnostic items, indicating at least some insight into the quality of the cues. Contrary to our expectations, experienced clinicians and novices had the same gaze patterns. Likewise, the two groups did not differ in the correspondence between the diagnostic gaze proportion and the quality of their responses. In conclusion, our results did not support the hypothesis that experienced clinicians' decision making is guided by more experience-based, encapsulated processes than is novices'.

What could be reasons for such a result? First, we turn to accuracy: it is possible that the more experienced clinicians felt pressure to be very accurate in their choices, which caused them to repeat steps, leading to slower overall search performance (Andersson, 2004). This interpretation is in line with previous results (Horstmann, Ahlgrimm, & Glöckner, 2009; Huang & Kuo, 2011) showing that when participants were told to be as accurate as possible, their eye-tracking data resulted in longer fixations. Another reason why more experienced clinicians were slower could be that they were older. Age and experience were confounded in our sample with a correlation of $r(27) = 0.90$, $p = 0.001$; thus, longer decision times in more experienced, that is, older participants, might simply reflect longer processing times and potentially a decline in working memory (Peters, Hess, Västfjäll, & Auman, 2007).

Second, our task setup might contribute to the divergence from results in the literature on choices and dwell differences. Our analysis of choices and dwell time was motivated by two related results: (i) the finding that items that have been looked at for longer also have a higher probability of ultimately being chosen (Krajbich, Armel, & Rangel, 2010); and (ii) the finding of an increase in gaze length at the chosen option right before choice (gaze cascade effect; Shimojo, Simion, Shimojo, & Scheier, 2003). We did not replicate these finding; one explanation could be that, in our task, eight cues pointed to either of two diagnoses (Figure S1). Participants thus looked at cues before they finally looked at their ultimately chosen option. In the previously cited studies, but not in ours, the to-be-fixated options are identical with the to-be-chosen options. Such gaze biases and gaze cascades are much weaker in our setup with to-be-fixated cues and to-be-chosen options.

Outlook

First and foremost, we wish to emphasize that our data say little about the value of psychotherapy or about experience-related differences in the treatment process. Rather, our focus is on a specific subtask at the beginning of the clinical process: the diagnostic classification. We chose a presentation format that allowed us to use the eye-tracking method, realizing that, in practice, clinical psychologists will not be presented with symptoms on a computer screen but with a description of symptoms in verbal or text format. However, clinical psychologists do have to give their patients' symptoms a diagnostic label, which is required for health insurance purposes (e.g. a Diagnostic and Statistical Manual of Mental Disorders classification is required in the Netherlands; ZorgWijzer.nl, 2014). It therefore seems relevant that practising clinicians are able to perform this decision task well. Research has also shown that presenting short clinical vignettes is a valid method for measuring conclusions in health assessments (Peabody, Luck, Glassman, Dresselhaus, & Lee, 2000). Of course, one may doubt the validity of the diagnostic labels themselves (e.g. Frances, 2013; Frances & Widiger, 2012), but such a discussion is beyond the scope of our study.

With increasing pressure on healthcare budgets, clinical psychologists are being held more and more accountable for their judgments—for their diagnostic classifications as well as for their assessment and treatment decisions (Wood, Garb, Lilienfeld, & Nezworski, 2002). They are increasingly expected to be scientists as well as practitioners, and their assessments are expected to be reliable, valid and to have proven treatment utility (Nelson-Gray, 2003). As our results show, there is certainly room for improvement here (Garb, 2005; Tracey *et al*., 2014; Vollmer, Spada, Caspar, & Burri, 2013).

Tracey and colleagues (2014) concluded that psychotherapy is a profession without any expertise, as therapists do not seem able to benefit from experience. We do not lay the blame on clinicians but explain this finding by stressing the difficulties inherent in the profession: having to work with

DOI: 10.1002/bdm

dynamic stimuli, lacking predictability and with no or ambiguous feedback to learn from (compared with Shanteau, 1992; Weiss & Shanteau, 2004). The clinical psychology domain differs from other domains such as weather forecasting or insurance analysis in allowing only limited competence because of its changeable properties with an inter-judge agreement of no more than 0.40 (Shanteau, 2000). The possibility of developing expertise is thus, at least partly, domain dependent and, in clinical psychology, hard to accomplish.

Our findings that consistency of diagnostic quality is low and that calibration varies considerably on the individual level independent of expertise also indicate that a focus on individual differences is warranted. Indeed, it might be advisable to focus on individual differences between clinicians rather than on years of experience, as proposed by Shanteau and Weiss (2014; Witteman, Weiss, & Metzmacher, 2012).

We believe that eye-tracking data offer insight into predecisional information search processes, allowing research questions that go beyond simple measures of outcome accuracy or thinking aloud data. In this study of clinical psychologists' diagnostic decision making, they unfortunately did not help producing findings that cast light on the expertise of clinical psychologists, which thus far still remains invisible.

ACKNOWLEDGEMENTS

We thank the incredible Sussanah Goss for editing this paper and for being the awesome person she is.

REFERENCES

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S.,… Rush, J. D. (2006). The meta-analysis of clinical judgment project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, *34*, 341–382. doi:10.1177/0011000005285875

Andersson, P. (2004). Does experience matter in lending? A process-tracing study on experienced loan officers' and novices' decision behavior. *Journal of Economic Psychology*, *25*, 471–492. doi:10.1016/S0167-4870(03)00030-8.

Ashby, N. J. S., Dickert, S., & Glöckner, A. (2012). Focusing on what you own: Biased information uptake due to ownership. *Judgment and Decision Making*, *7*, 254–267.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from http://CRAN.R-project.org

Brailey, K., Vasterling, J. J., & Franks, J. J. (2001). Memory of psychodiagnostic information: Biases and effects of expertise. *American Journal of Psychology*, *114*, 55–92. doi:10.2307/1423381.

Brammer, R. (2002). Effects of experience and training on diagnostic accuracy. *Psychological Assessment*, *14*, 110–113. doi:10.1037/1040-3590.14.1.110.

Croskerry, P., & Norman, G. (2008). Overconfidence in clinical decision making. *The American Journal of Medicine*, *121*, S24–S29. doi:10.1016/j.amjmed.2008.02.001.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, *85*, 395–416. doi:10.1037/0033-295X.85.5.395.

Elstein, A. S., & Schwartz, A. (2002). Evidence base of clinical diagnosis. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *British Medical Journal*, *324*, 729–732. doi:10.1136/bmj.324.7339.729.

Frances, A. (2013). The new somatic symptom disorder in DSM-5 risks mislabeling many people as mentally ill. *British Medical Journal*, *346*, f1580. doi:10.1136/bmj.f1580.

Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, *8*, 109–130. doi:10.1146/annurev-clinpsy-032511-143102.

Friedlander, M. L., & Phillips, S. D. (1984). Preventing anchoring errors in clinical judgment. *Journal of Consulting and Clinical Psychology*, *52*, 366–371. doi:10.1037/0022-006X.52.3.366.

Garb, H. N. (1986). The appropriateness of confidence ratings in clinical judgment. *Journal of Clinical Psychology*, *42*, 190–197. doi:10.1002/1097-4679.

Garb, H. N. (1998). *Studying the clinician: Judgment research and psychological assessment*. Washington, DC: American Psychological Association.

Garb, H. N. (2005). Clinical judgment and decision making. *Annual Review of Clinical Psychology*, *1*, 67–89. doi:10.1146/annurev.clinpsy.1.102803.143810.

Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational Psychology Review*, *23*, 523–552. doi:10.1007/s10648-011-9174-7.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528. 10.1037/0033-295X.98.4.506.

Hogarth, R. M. (2010). Intuition: A challenge for psychological research on decision making. *Psychological Inquiry*, *21*, 338–353. doi:10.1080/1047840X.2010.520260.

Horstmann, N., Ahlgrimm, A., & Glöckner, A. (2009). How distinct are intuition and deliberation? An eye-tracking analysis of instruction-induced decision modes. *Judgment and Decision Making*, *4*, 335–354. doi:10.2139/ssrn.1393729.

Huang, M. Y., & Kuo, F. (2011). An eye-tracking investigation of internet consumers' decision deliberateness. *Internet Research*, *21*, 541–561. doi:10.1108/10662241111176362.

Inquisit 3.0.4.0 [Computer software]. (2009). Seattle, WA: Millisecond Software LLC.

Kim, N. S., & Ahn, W. (2002). Clinical psychologists' theory-based representations of mental disorders predict their diagnostic reasoning and memory. *Journal of Experimental Psychology: General*, *131*, 451–476. doi:10.1037/0096-3445.131.4.451.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Publishing Group*, *13*, 1292–1298. doi:10.1038/nn.2635.

Menkhoff, L., Schmeling, M., & Schmidt, E. (2013). Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization*, *86*, 92–101. doi:10.1016/j.jebo.2012.12.022.

Nelson-Gray, R. O. (2003). Treatment utility of psychological assessment. *Psychological Assessment*, *15*, 521–531. doi:10.1037/1040-3590.15.4.521.

Peabody, J. W., Luck, J., Glassman, P., Dresselhaus, T. R., & Lee, M. (2000). Comparison of vignettes, standardized patients, and chart abstraction: A prospective validation study of 3 methods for measuring quality. *JAMA*, *283*, 1715–1722. doi:10.1001/jama.283.13.1715.

Peters, E., Hess, T. M., Västfjäll, D., & Auman, C. (2007). Adult age differences in dual information processes. Implications for the role of affective and deliberative processes in older adults' decision making. *Perspectives on Psychological Science*, *2*, 1–23. doi:10.1111/j.1745-6916.2007.00025.x.

Ranyard, R., & Svenson, O. (2011). Verbal data and decision process analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & R. Ranyard (Eds.), *A Handbook of Process Tracing Methods for Decision Research* (pp. 115–138). Psychology Press.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*, 759–769.

DOI: 10.1002/bdm

Schmidt, H. G., & Rikers, R. M. J. (2007). How expertise develops in medicine: Knowledge encapsulation and illness script formation. *Medical Education*, *41*, 1133–1139. doi:10.1111/j.1365-2923.2007.02915.x.

Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011a). *A handbook of process tracing methods for decision research*. Oxford, United Kingdom: Taylor & Francis.

Schulte-Mecklenbeck, M., Kühberger, A., & Ranyard, R. (2011b). The role of process data in the development and testing of process models of judgment and decision making. *Judgment and Decision Making*, *6*, 733–739.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, *53*, 252–266. doi:10.1016/0749-5978(92)90064-E.

Shanteau, J. (2000). Why do experts disagree?. In Green, B., et al. (Eds.), *Risk behaviour and risk management in business life* (pp. 186–196). Dordrecht, The Netherlands: Kluwer Academic Press.

Shanteau, J., & Weiss, D. J. (2014). Individual expertise versus domain expertise. *American Psychologist*, *69*, 711–712. doi:10.1037/a0037874.

Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, *6*(12), 1317–1322. doi:10.1038/nn1150.

Spengler, P. M., White, M. J., Ægisdóttir, S., Maugherman, A. S., Anderson, L. A., Cook, R. S.,… Rush, J. D. (2007). The meta-analysis of clinical judgment project: Effects of experience on judgment accuracy. *The Counseling Psychologist*, *37*(3), 350–399. doi:10.1177/0011000006295149

Strasser, J., & Gruber, H. (2004). The role of experience in professional training and development of psychological counsellors. In Boshuizen, H. P. A., Bromme, R., & Gruber, H. (Eds.), *Professional learning: Gaps and transitions on the way from novice to expert* (pp. 11–27). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Tracey, T. J. G., Wampold, B. E., Lichtenberg, J. W., & Goodyear, R. K. (2014). Expertise in psychotherapy: An elusive goal? *American Psychologist*, *69*, 218–229. doi:10.1037/a0035099.

Vollmer, S., Spada, H., Caspar, F., & Burri, S. (2013). Expertise in clinical psychology. The effects of university training and practical experience on expertise in clinical psychology. *Frontiers in Psychology*, *4*, 1–12. doi:10.3389/fpsyg.2013.00141.

Weiss, D. J., & Shanteau, J. (2004). The vice of consensus and the virtue of consistency. In Smith, K., Shanteau, J., & Johnson, P. (Eds.), *Psychological investigations of competence in decision making* (pp. 226–240). Cambridge, UK: Cambridge University Press.

Witteman, C. L. M., & Tollenaar, M. S. (2012). Remembering and diagnosing clients: Does experience matter? *Memory*, *20*, 266–276. doi:10.1080/09658211.2012.654799.

Witteman, C. L. M., & Van den Bercken, J. H. L. (2007). Intermediate effects in psychodiagnostic classification. *European Journal of Psychological Assessment*, *23*, 56–61. doi:10.1027/1015-5759.23.1.56.

Witteman, C. L. M., Weiss, D. J., & Metzmacher, M. (2012). Assessing diagnostic expertise of counselors using the Cochran–Weiss–Shanteau (CWS) index. *Journal of Counseling and Development*, *90*, 30–24. doi:10.1111/j.1556-6676.2012.00005.x.

Wood, J. M., Garb, H. N., Lilienfeld, S. O., & Nezworski, M. T. (2002). Clinical assessment. *Annual Review of Psychology*, *53*, 519–543. doi:10.1146/annurev.psych.53.100901.135136.

ZorgWijzer.nl (2014). Psychologische zorg [Psychological care]. Retrieved November 2014 from http://www.zorgwijzer.nl/vergoeding/psychologie

Authors' biographies:

Michael Schulte-Mecklenbeck is a lecturer in Consumer Behavior at the University of Bern and an Adjunct Researcher at the Max Planck Institute for Human Development, Berlin. He received his PhD from the University of Fribourg, Switzerland in 1998, worked in industrial consumer research at a larger food company and is on a journey back to academia. His research focuses on cognitive processes in decision making, consumer behavior and food choice.

Nanon L. Spaanjaars is a healthcare psychologist working at GGNet, a large institute for mental health in the Netherlands. She completed the post-graduate professional training programme for health care psychologist in 2015. Previously, she received both her Clinical Master (in 2011) and her Research Master Behavioural Science (in 2009) from Radboud University, Nijmegen, the Netherlands.

Cilia L.M. Witteman is full professor of Psychodiagnostic decision making at the Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands. She is past president of the European Association for Decision Making. Her PhD in 1992 is from Utrecht University, the Netherlands, on Belief Revision. Her research is on judgment and decision making applied to mental health professionals and forensic situations, and also addresses the development of intuition.

Authors' addresses:

Michael Schulte-Mecklenbeck, Department of Business Administration, University of Bern, Switzerland; Max Planck Institute for Human Development, Berlin, Germany

Nanon L. Spaanjaars and Cilia L. M. Wittema, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

DOI: 10.1002/bdm