

# Tracks to a Medical Diagnosis: Expertise Differences in Visual Problem Solving

Citation for published version (APA):

Jaarsma, T., Boshuizen, H. P. A., Jarodzka, H., Nap, M., Verboon, P., & van Merriënboer, J. J. G. (2016). Tracks to a Medical Diagnosis: Expertise Differences in Visual Problem Solving. *Applied Cognitive Psychology*, 30(3), 314-322. <https://doi.org/10.1002/acp.3201>

## Document status and date:

Published: 01/01/2016

## DOI:

[10.1002/acp.3201](https://doi.org/10.1002/acp.3201)

## Document Version:

Publisher's PDF, also known as Version of record

## Document license:

Taverne

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

## Tracks to a Medical Diagnosis: Expertise Differences in Visual Problem Solving

THOMAS JAARSMA<sup>1\*</sup>, HENNY P. A. BOSHUIZEN<sup>1,2</sup>, HALSZKA JARODZKA<sup>1,3</sup>, MARIUS NAP<sup>4</sup>,  
PETER VERBOON<sup>1,6</sup> and JEROEN J. G. VAN MERRIËNBOER<sup>1,5</sup>

<sup>1</sup>*Welten Institute, Open University of The Netherlands, Heerlen, The Netherlands*

<sup>2</sup>*University of Turku, Turku, Finland*

<sup>3</sup>*Humanities Laboratory, Lund University, Lund, Sweden*

<sup>4</sup>*Pathology Department, Atrium Medical Center, Heerlen, The Netherlands*

<sup>5</sup>*Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands*

<sup>6</sup>*Faculty of Psychology and Educational Sciences, Open University of The Netherlands, Heerlen, The Netherlands*

*Summary: This study focuses on the visual problem-solving process of clinical pathologists. Its aim is to find expertise-related differences in the temporal arrangement of this process, with a special focus on the orientation phase. A theoretical model of the visual diagnostic process of medical specialists is extended with general problem-solving theory. Participants were 13 experts, 12 intermediates, and 13 novices, who all diagnosed seven microscopic images. Their microscope movements and thinking aloud were recorded. To study temporal arrangement of the process, we applied a time-grid to the data. The results reflected several aspects of general problem-solving theory. Experts and intermediates showed a more extensive orientation phase and more refined schemata than novices. Intermediates also showed a control phase at the end of the diagnostic process. Novices showed a uniform process. These phases were reflected in microscope navigation and thinking aloud, which justifies the extension of the theoretical model. Copyright © 2016 John Wiley & Sons, Ltd.*

### INTRODUCTION

Visual problem solving is an important component of the work of professionals such as air traffic controllers (Van Meeuwen et al., 2014) or medical specialists, for instance radiologists or pathologists (Reingold & Sheridan, 2011). All these professionals solve problems through the interpretation of complex, information-dense visualisations: They interpret visual information and process this information cognitively. Newcomers in these professions are often trained on the job by their expert colleagues, lacking a clear curriculum or didactic plan. As a first step towards such a didactic plan, insight into the developmental stages that novices and intermediates go through to become experts, ought to be obtained. This insight could improve feedback during training situations, and thus the training itself (Shute & Towle, 2003). The aim of this study therefore is to gain insight into the developmental stages of visual problem-solving expertise in a complex, knowledge-intensive and highly visual domain: clinical pathology—a domain in which medical specialists diagnose tissue samples microscopically.

Literature does not provide a clear view on the development of visual problem-solving expertise. The closest it gets may be the ‘visual search and detection model’ by Nodine and Kundel (1987). This model describes the diagnostic process of expert radiologists and divides it into two phases based on two different modes of viewing. The first phase is characterised by holistic viewing. In this phase, a global impression of the image is obtained by means of pattern recognition and quickly compared with mental images of normal anatomy, which allows the diagnostician to determine perturbations of the normal situation. These abnormalities are then focally examined in the second phase. As a possible

additional step, the diagnostician might apply a ‘stereotyped scanning pattern’ to see if there is any new, contrasting information in the image (Nodine & Kundel, 1987, p. 1246).

There are several reasons that this model does not suffice to describe visual problem solving. First of all, it only describes visual processes: It is purely based on eye tracking data, ignoring the cognitive processing of visual information. Second, the model has been based on a specific domain (radiology) and on a specific task in this domain, namely, the detection of abnormalities (nodules) in static X-ray images. These are—for experts—highly automatised detection tasks, predominantly relying on visual perception. This corresponds with the very brief first phase of holistic viewing, which lasts ‘a few hundred milliseconds’ (Nodine & Kundel, 1987, p. 1246), or 2 seconds at most (Mello-Thoms et al., 2005). Such a task is essentially a ‘visual-search problem’ (Nodine & Mello-Thoms, 2010, p. 142), rather than visual problem solving.

The problems central in visual problem solving, such as those that clinical pathologists solve, are different in more than one aspect. First of all, the microscopic images in clinical pathology are interactive in the sense that they typically need manipulation to be optimally informative (through zooming and panning). The information necessary for a diagnosis thus needs to be searched for manually. This puts the diagnostic process in clinical pathology on a different timescale than the very brief visual-search tasks on which the model by Nodine and Kundel (1987) is based. Besides, diagnostic tasks in clinical pathology (as in radiology) are often not simple detection tasks: Patterns and single features need to be interpreted in combination with background information of the patient to come to a diagnosis. Such an interpretation task is likely to demand more cognitive processing than a feature detection task, which is indeed highly visual. We argue that the diagnostic task in clinical pathology should thus be seen as a case of visual problem solving, rather than a visual-search problem. Consequently, we thus

\* Correspondence to: Thomas Jaarsma, Faculty of Social and Behavioural Sciences, University of Amsterdam, PO Box 15780, 1001 NG Amsterdam, The Netherlands.  
E-mail: t.jaarsma@uva.nl

need to expand the visual model of the diagnostic process with a theory on problem solving.

To include general problem-solving processes, we draw on the early work on problem solving of Newell and Simon (1972). They identified two phases in problem solving: (i) the construction of a problem representation and (ii) the search for a solution within that problem representation. When confronted with a problem, the solver constructs a problem representation through the activation of schemata in long-term memory (Chi & Glaser, 1985). As the solver's internal representation of the task, it includes a problem statement, the solver's relevant knowledge, and the possible strategies to solve the problem (Newell & Simon, 1972). The problem representation thus plays an important role in problem solving, as it determines the playing field for the solution process. And what is more important, the appropriateness and quality of this playing field are directly influenced by the availability of specific and appropriate schemata. Clinical pathologists, as all other problem solvers, thus need adequate knowledge structures to correctly interpret microscopic images.

Boshuizen and Schmidt (1992) have studied the differences in the knowledge structures between medical specialists with different levels of expertise. Novices typically possess detailed, extensive knowledge networks, consisting mainly of biomedical concepts. The knowledge networks of more experienced diagnosticians include knowledge encapsulations: subsumptions of 'lower level, detailed propositions under higher level, sometimes clinical propositions' (Boshuizen & Schmidt, p. 168). For expert problem solvers, the cognitive structure takes on a script or schema type of knowledge organisation. These schemata include components such as enabling conditions, the pathophysiological process, and the symptoms resulting from that process (Boshuizen & Schmidt, 2008). When confronted with a diagnostic task, high-expertise diagnosticians are thus likely to activate more advanced schemata, which are directly indicative for the diagnosis, whereas low-expertise diagnosticians come up with detailed knowledge concepts. Typically, experts take a larger proportion of their time-on-task to construct such a thorough representation (Chi, Glaser, & Rees, 1981; Crowley, Naus, Stewart, & Friedman, 2003; Van Gog, Paas, & Van Merriënboer, 2005).

A study in the visual medical domain that applies the principle of schemata is the study on radiologists by Lesgold et al. (1988). Its focus is on how radiologists with different levels of expertise activate, test, and consequently refute or refine their schemata when diagnosing X-ray images. The results show that experts activate more appropriate schemata than intermediates and do so earlier in the diagnostic process. After instantiation, experts apply better methods to test their schemata and show more flexibility towards them than novices. Whereas novices may be tempted to interpret abnormalities in accordance with their activated schemata, experts know when new information requires adjustment of the schemata. Also, experts are opportunistic: Valuable information that suddenly pops up is valued and can lead to tuning or rejection of a schema. This study thus shows that a more cognitive approach to the visual diagnostic process renders valuable insights.

The theory of general problem solving thus discerns different phases in the problem-solving process and associates the construction of an orientation phase with expertise. To test the applicability of this theory to visual problem solving in clinical pathology, we thus need to be able to determine the presence and nature of the orientation phase. Put differently, we need to be able to differentiate between several parts of the process. The application of a time-grid offers the opportunity to do so. Instead of considering the diagnostic process as a whole, an analysis with a time-grid divides the process into a fixed number of segments, equal in length. By comparing these segments, we are able to examine the temporal arrangement of the diagnostic process: We can determine the presence and nature of an orientation phase, and any other kind of phasing occurring in the diagnostic processes of our participants. To cover the full diagnostic process, both visual and cognitive aspects need to be included: Microscope movements will be registered to study visual processes, while concurrent think aloud data will be collected on the cognitive processes.

The following research questions are formulated for this study:

1. Which expertise-related differences can be found in the temporal arrangement of the diagnostic process of clinical pathologists, as expressed in microscope navigation and clinical reasoning?
2. How is an orientation phase reflected in the temporal arrangement of microscope navigation and clinical reasoning, and which expertise-related differences can be identified therein?
3. How do the results of this study inform a new, more encompassing model of the visual problem-solving process of medical specialists?

## METHODS

### Participants and design

The participants in this study ( $N=38$ ,  $M=35.39$  years,  $SD=14.67$ ; 24 women) were recruited on a voluntary basis in two hospitals in the Netherlands. Based on their experience, they formed three expertise levels: experts were 13 clinical pathologists ( $M=51.77$  years,  $SD=10.05$ ; three women) with an average experience of 21.38 years ( $SD=10.03$ ), including 5 years of training. Intermediates were 12 residents ( $M=33.25$  years,  $SD=6.28$ ; eight women) with an average of 3 years of training ( $SD=1.60$ ). Novices were 13 second-year medical students who had completed two courses in the physiology and pathology of cells and tissue, including the gastrointestinal tract ( $M=21.00$  years,  $SD=2.58$ , all female). All participants had normal or corrected to normal eyesight. They received a small gift for their participation (book voucher) after the experiment.

The experiment was set up as a between-subjects design, with expertise level and segment as the independent variables. The dependent variables were drawn from microscope movements and thinking aloud; they are specified in the Data Reduction and Analysis section. The first case was the same for all participants (allowing us to delete just one case if

participants showed great effort to get used to the virtual microscope—which was not the case), while the remaining six were arranged according to a balanced Latin square. The study was approved by the ethical committee of the Open University, who did not deem it to require any kind of formal medical-ethical approval because no patients were involved. All participants gave written informed consent.

## Materials and apparatus

### Cases

The seven cases in the experiment all involved microscopic images of the colon, which were obtained from the Atrium Medical Centre in Heerlen, the Netherlands. *A priori*, these cases were diagnosed by four clinical pathologists, and a consensus diagnosis was reached for all cases. The cases were divided over five diagnostic categories: adenocarcinoma (cancer, two cases, one of them being the first case for all participants), adenoma (pre-stage of cancer, two cases), inflammation (one case), hyperplastic polyp (benign polyp, one case), and normal tissue (one case).

To be able to record the participants' zooming and panning movements, the Aperio ImageScope digital microscope (version 11.2.0.780) of Leica Biosystems (Wetzlar, Germany) was used, in combination with a 22-in. digital monitor (model P2210 by Dell, Round Rock, TX) with a resolution of 1680 × 1050 pixels. Using a digital microscope that resembles navigating through a digital map, one can zoom in and zoom out and pan in the horizontal plane. All these movements were recorded by the digital microscope for later analysis. The participants were not acquainted with the virtual microscope prior to taking part in the experiment: They either used light microscopes (experts and intermediates) or were generally unexperienced with microscopy (novices).

## Procedure

The experiment was carried out in individual sessions of about 30 minutes. First, participants completed the demographic questionnaire. Then, they were introduced to the digital microscope and solved a practice case (not included in the data collection) to become acquainted with the microscope operation.

The setting simulated the normal working procedure. For all cases, a written patient background was presented first, including sex, age, kind of tissue, and comments or questions from the requesting physician. The latter comments differed in how informative they were: They varied from a likely diagnostic category, a description of the material, or a specific question. Then, participants entered the digital microscope and started the diagnostic process. During the entire diagnostic process (i.e. from opening the image to closing it), they thought aloud. As soon as they had come to a diagnosis, the participants closed the slide and pushed a button to proceed to selecting one out of five diagnostic categories: normal, adenoma, adenocarcinoma, inflammation, or hyperplastic polyp. This procedure was repeated for each case. The viewing duration per case (i.e. the time-on-task) was on average little under 2 minutes, with novices ( $M = 152$  seconds,  $SD = 18.91$ ), taking more time than both

intermediates ( $M = 110$  seconds,  $SD = 19.62$ ) and experts ( $M = 86$  seconds,  $SD = 18.91$ ).

## Data reduction and analysis

The experiment resulted in a total of 261 records of microscope navigation and 266 verbal protocols of the participants' thinking aloud (seven cases times 38 participants; the recordings of the microscope navigation of five trials were lost owing to technical problems).

### Microscope navigation

The data recorded by the digital microscope consisted of the coordinates (in pixels) of the image sections that had been viewed by each participant, at which magnification and for how long. Accidental zooming actions (i.e. zooming actions followed within 1 second by a movement in the opposite direction) were deleted from the data.

Next, a time-grid was applied to these files. Each file was divided into four segments of equal length, based on the individual time-on-task (time from opening the case until closing it). Subsequently, the microscope movements were assigned to a segment. As an example, a fictitious case that would have lasted 2 minutes would have been divided into four segments for 30 seconds. Those movements that ended during the first 30 seconds were assigned to the first segment, those that ended between 30 and 60 seconds to the second, and so on. Seven measures were calculated per segment: *average magnification* (a time-weighted average of magnification), *minimal magnification*, *maximal magnification*, *number of panning movements*, *number of opposed zooming movements* (change in zooming direction after two consecutive movements in the same direction), *hits of relevant areas* (number of cases when the centre of the screen fell into a diagnostically relevant area, designated as such by an expert pathologist), and *time spent in relevant areas*.

The measures per segment were statistically analysed in two steps. To account for possible correlations between measures, a multivariate analysis of variance (MANOVA) was carried out first. In this MANOVA, expertise level and segment were the independent variables, while all measures described earlier were included as dependent variables. Those measures with a *p*-value below .10 on the interaction effect of expertise level and segment were subsequently analysed separately in a generalised linear mixed model. This latter model took the skew distribution of the data into account, as well as the hierarchy in the data (a multilevel model with measurements on the first level, and subjects and cases on the second).

### Clinical reasoning

Per-case thinking aloud data were collected, adding up to a total of 266 verbal protocols. The audio recordings were transcribed by a research assistant and checked for correctness and completeness by the first author. A primarily bottom-up approach was used to design a coding scheme. In the first stage of this process, the first author read protocols, drafted initial codes, and tested these on a small subset of transcripts. This cycle was repeated a few times with a set of codes as the outcome. Then, this set was compared and

aligned with the coding scheme designed by Crowley et al. (2003), forming a tentative coding scheme. Utterances were based on content and differed in length: A finding could be one word, while a reflection on one's performance could be a full sentence. This tentative scheme was tested and fine tuned in a series of inter-rater reliability tests on a subset of the protocols with the first and second authors. This resulted in a coding scheme of 26 main categories, with a percent agreement of 72.4% between the two raters (Krippendorff's  $\alpha$  of 0.67). The first author then coded the remaining protocols. The intra-rater reliability was 91% (Krippendorff's  $\alpha$  of 0.89). See Table 1 for definitions and examples of the codes.

The allocation to a segment of the diagnostic process was based on the sequence of utterances. Per protocol, each utterance was given a sequence number: The first utterance was given number 1, the second number 2, and so on. The total number of utterances was divided by four, to determine four segments. Then, each utterance was assigned to a segment based on its sequence number. So if there were 20 coded utterances in one protocol, utterances 1–5 were put in Segment 1, 5–10 in Segment 2, and so on.

To explore the relations between utterances, segments and expertise level the verbal data were analysed by multiple correspondence analysis (MCA). In MCA (Gifi, 1990), the aim is to visualise the relationships between the several categorical variables in a low-dimensional space (two or three dimensions). The two dimensions represent the main discriminators: They describe the central oppositions in the data. People who fall in categories far away from the origin have distinct scores on the other variable, whereas people in categories that are positioned close to the origin resemble the average pattern of scores on the other variable.

## RESULTS

### Microscope navigation

#### *multivariate analysis of variance*

Using Pillai's trace, there was a significant effect of expertise level ( $V=0.22$ ,  $F(14, 1448)=12.50$ ,  $p < .01$ ,  $\eta_p^2=0.11$ ), segment ( $V=0.58$ ,  $F(21, 2175)=24.97$ ,  $p < .01$ ,  $\eta_p^2=0.19$ ), and the interaction expertise level\*segment ( $V=0.14$ ,  $F(42, 4368)=12.50$ ,  $p < .01$ ,  $\eta_p^2=0.02$ ) on the group of seven

Table 1. Categories of clinical reasoning, including definitions and examples

Main categories	Definition	Example
1. Patient background	Patient background information	<i>So this is a 76-year old woman.</i>
2. Quantity and nature of material	Counting pieces or identifying colon	<i>I see multiple fragments of colon tissue.</i>
3. Spatial orientation of tissue pieces	Mental rotation of pieces to get an overview of the material	<i>This piece here is the same as this, but rotated.</i>
4. Artefact in image	Artefacts caused by scanning of the tissue	<i>This scan is out of focus.</i>
5. Artefact in tissue	Artefacts caused by preparation of tissue	<i>There is thermic damage to the tissue here.</i>
6. Findings—colour and shape	Findings described by colour, size, and/or shape	<i>There is a pink stain here. This looks like a flower.</i>
7. Findings—impression	Findings phrased as an impression	<i>This looks weird.</i>
8. Findings—name	Specific abnormalities	<i>Here is necrosis.</i>
9. Findings—architecture	Regarding the spatial arrangement of cells	<i>Neatly aligned crypts.</i>
10. Comparison between cases	Comparisons with previous cases of experiment	<i>There are more crypts here than in the previous one. It is darker here.</i>
11. Comparison within case	Identification of current finding through comparison with previous one	<i>Same here. Here also.</i>
12. Repetition within case	Repetition of previous findings	<i>The crypts look typical.</i>
13. Comparison with normal tissue	Implicit comparisons to mental images of normal tissue	<i>Normal structure of the epithelium. I do not see any inflammation. So this could be Crohn.</i>
14. Absent features	Specific features reported as lacking	<i>Yes, carcinoma.</i>
15. Hypothesis	Hypotheses on the diagnosis	<i>It's not a hyperplastic polyp.</i>
16. Diagnosis	Comments with a diagnosis	<i>Not malign, anyway.</i>
17. Exclusion of diagnosis	Exclusions of a diagnosis	<i>I zoom in here.</i>
18. Rejection of hypothesis	Rejection of a hypothesis	<i>Let's see if I can find that back here.</i>
19. Heuristic comment	References to upcoming actions	<i>I am going to look for a desmoplastic reaction.</i>
20. Search goal	Specific search objectives	<i>I really don't know anything!</i>
21. Reflection—knowledge and skills	Reflections on own knowledge and skills	<i>So the diagnosis primarily depends on the clinical information.</i>
22. Reflection—diagnosis	Anything on the diagnostic process as such	<i>I would not have given a diagnosis here.</i>
23. Reflection—normal process	Comments on what the participant would have done in reality	<i>This is not a very neat polyp.</i>
24. Reflection—solvability of slide	Comments on how high the solvability of the slide is	<i>I probably need to diagnose it anyway.</i>
25. Reflection—experiment	Utterances on the experimental procedure	<i>Because it continues below this layer.</i>
26. Knowledge statements and inferences	Comments that contain knowledge or reasoning	<i>I think a polyp looks different.</i>

dependent variables of microscope navigation (average magnification, minimal magnification, maximal magnification, panning movements, opposed zooming movements, hits of relevant areas, and time spent in relevant areas). Separate ANOVAs on these dependent variables revealed significant interaction effects of expertise level \* segment on the average magnification ( $F(6, 729)=4.08, p < .01, \eta_p^2=0.03$ ), minimal magnification ( $F(6, 729)=4.24, p < .01, \eta_p^2=0.03$ ), maximal magnification ( $F(6, 729)=3.19, p < .01, \eta_p^2=0.03$ ), and opposed zooming movements ( $F(6, 729)=3.49, p < .01, \eta_p^2=0.03$ ). There was a marginally significant interaction effect on hits of relevant areas ( $F(6, 729)=1.86, p = .09, \eta_p^2=0.02$ ). As the interaction effects are of primary interest to this study, we will focus on them in our further analysis.

#### Generalised linear mixed model

For the descriptive statistics of the interaction effect between expertise level and segment, please see Table 2.

Univariate analyses using the generalised linear mixed model revealed the same significant interaction effects as did the MANOVA: average magnification ( $F(6, 1028)=2.29, p = .03$ ), minimal magnification ( $F(6, 1028)=2.86, p = .01$ ), maximal magnification ( $F(6, 1028)=2.30, p = .03$ ), and the number of opposed zooming movements ( $F(6, 1028)=4.51, p < .01$ ). There was a marginally significant interaction effect on hits of relevant areas ( $F(6, 730)=1.95, p = .07$ ).

For all expertise levels, the average magnification in the first segment was significantly lower than in the other segments. For novices and experts, Segments 2–4 were equal, while for intermediates, the average magnification of Segment 4 was lower than that of Segment 3. Intermediates, thus, had a distinct middle part with a higher average magnification. This same pattern is present in the minimal magnification, but not in the maximal magnification. For novices,

maximum magnification is equally high for Segments 1–3, while it is lower in Segment 4. For experts and intermediates, Segment 1 contrasts with Segments 2–4, having a lower maximal magnification.

For the number of opposed zooming movements, intermediates and experts showed an inactive first segment. The most interesting difference for this variable lies in Segment 4, where the intermediates' number of opposed zooming movements was higher than that of the other segments, while for experts, it remained the same. Novices showed no differences between segments for this variable.

Finally, experts were the only group with differences between segments for the hits of relevant areas: In Segment 1, they hit these relevant areas less often than in the later segments. For intermediates and novices, this number was equal throughout all segments. This indicates that intermediates and novices equally focus on relevant areas from the beginning onwards, whereas experts do so from Segment 2 onwards.

#### Clinical reasoning

The two dimensions identified by the MCA accounted for 94% of the variance (48% by dimension 1 and 46% by dimension 2). Figure 1 displays the two-dimensional space with the coding categories, the segments, and the expertise levels jointly represented. Dimension 1 represents differences between the four segments, distinguishing Segment 1 from Segment 4 with Segments 2 and 3 close together in the middle. Dimension 2 is predominantly characterised by differences between expertise levels, distinguishing novices versus intermediates and experts. Table 3 gives the frequencies of the codes, split up per expertise level, and segment.

As appears from Figure 1, the most pronounced temporal differences between coding categories are between Segments 1 and 4. Utterances on patient background (code 1) and on

Table 2. Descriptive statistics [mean (SD)] for all variables of microscope navigation with a significant interaction effect of segment and expertise level

		Segment 1	Segment 2	Segment 3	Segment 4	Total
Novices	Average magnification	4.52 (0.41)	5.92 (0.52)	6.53 (0.57)	5.58 (0.49)	5.63 (0.41)
	Minimal magnification	1.33 (0.14)	3.94 (0.31)	3.73 (0.30)	3.64 (0.29)	3.16 (0.19)
	Maximal magnification	8.42 (0.95)	9.11 (1.02)	9.52 (1.07)	7.75 (0.88)	8.70 (0.82)
	Opposed zooming movements	1.27 (0.26)	1.40 (0.28)	1.55 (0.31)	1.53 (0.30)	1.44 (0.24)
	Hits relevant areas	7.69 (1.36)	7.64 (1.35)	7.55 (1.33)	6.96 (1.23)	7.47 (0.99)
Intermediates	Average magnification	3.92 (0.38)	6.80 (0.61)	6.83 (0.62)	5.76 (0.53)	5.82 (0.44)
	Minimal magnification	1.38 (0.15)	3.64 (0.31)	3.51 (0.30)	2.99 (0.22)	2.71 (0.18)
	Maximal magnification	7.20 (0.86)	10.10 (1.18)	10.44 (1.22)	9.17 (1.07)	9.22 (0.90)
	Opposed zooming movements	0.86 (0.19)	1.71 (0.35)	2.02 (0.41)	2.83 (0.56)	1.86 (0.32)
	Hits relevant areas	8.88 (1.64)	12.36 (2.26)	11.86 (2.17)	12.16 (2.22)	11.32 (1.56)
Experts	Average magnification	2.93 (0.29)	5.01 (0.45)	5.73 (0.51)	5.33 (0.48)	4.75 (0.35)
	Minimal magnification	1.48 (0.16)	3.26 (0.27)	3.76 (0.31)	3.06 (0.26)	2.89 (0.18)
	Maximal magnification	5.60 (0.65)	6.74 (0.77)	8.01 (0.92)	8.03 (0.91)	7.11 (0.67)
	Opposed zooming movements	0.36 (0.09)	0.67 (0.15)	1.23 (0.25)	1.34 (0.27)	0.90 (0.15)
	Hits relevant areas	3.38 (0.63)	5.70 (1.03)	6.72 (1.21)	7.34 (1.31)	5.79 (0.78)
Total	Average magnification	3.78 (0.21)	5.86 (0.30)	6.34 (0.33)	5.55 (0.29)	
	Minimal magnification	1.39 (0.09)	3.60 (0.17)	3.66 (0.17)	3.01 (0.15)	
	Maximal magnification	7.04 (0.47)	8.55 (0.56)	9.29 (0.61)	8.31 (0.55)	
	Opposed zooming movements	0.80 (0.10)	1.19 (0.14)	1.55 (0.18)	1.81 (0.21)	
	Hits relevant areas	4.79 (0.50)	7.29 (0.72)	6.77 (0.68)	7.12 (0.71)	

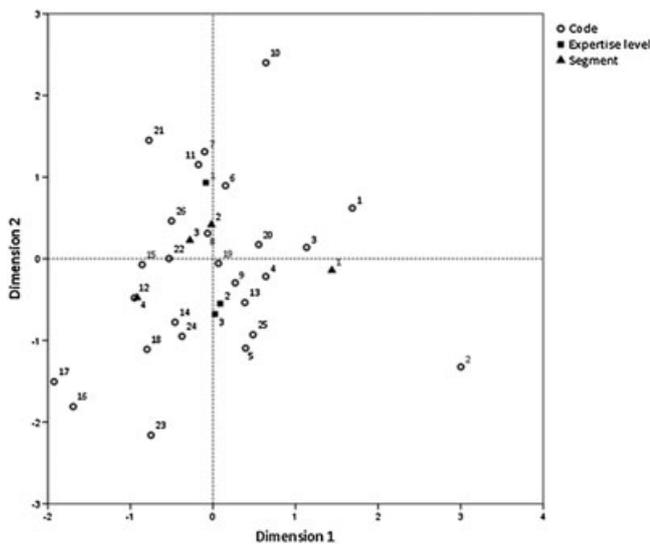


Figure 1. Visualization of the outcome of the multiple correspondence analysis (MCA) of the clinical reasoning. Dimension 1 contrasts the four segments (solid triangles, 1-4), dimension 2 the three expertise levels (solid squares, 1=*novices*; 2=*intermediates*, 3=*experts*). Coding categories are represented by the circles: the numbers correspond with those used in Tables 4.1 and 4.3. Segment 1 contributes most to dimension 1. Therefore, the higher a coding category scores on this dimension, the closer its association to segment 1 (e.g., category 2 – Quantity and nature of material). On the contrary, segment 4 is least associated with dimension 1: all categories on the left hand side are therefore associated to this segment. The same goes for dimension 2, which contrasts mainly between novices on the one hand, and intermediates and experts on the other

the quantity (2) and spatial orientation (3) of tissue pieces are associated with the first segment. Novices tended to say more about the patient background (1), while experts often

inventoried the number of tissue pieces on the slide (2). Segment 4 is associated with concluding and reflective comments: Intermediates and experts mentioned many diagnoses (16) and rejected hypotheses (18) in this segment. In addition, intermediates tended to exclude diagnoses (17) at this stage. Novices were less conclusive (16–18) and instead reflected on their knowledge and skills (21) and on the diagnosis (22). Experts also reflected in the fourth segment, but more on the diagnosis (22) and what they would have done normally (23).

The spread in all categories related to findings (6–9) is related to expertise level: Findings expressed by colour and shape (6), impressions (7), and names (8) are primarily associated with novices, whereas those expressed by architectural features (9) is slightly more present in intermediates and experts. Artefacts in the image (4) were associated with Segments 1 and 2. Artefacts in the tissue (5) are mentioned throughout the process, but mainly so by experts and intermediates. Novices are the only ones who verbalised comparisons between cases (10) and did so mainly at the start of the process (Segments 1 and 2). Novices not only compared between cases, but they also tended to compare within a case to interpret features (11). Experts often mentioned the repeated appearance of features (12) and did so mostly in Segment 4. As a final category of findings, absent features (14) were mentioned predominantly by intermediates and experts, and their frequency increased during the diagnostic process, from Segment 2 onwards.

There are a couple of process-oriented categories that revealed some subtle differences between segments and expertise levels. Search goals (20) were seldom stated in Segment 4 and were mentioned slightly less often by

Table 3. Frequencies of codes, split up per expertise level and segment

	Segment 1				Segment 2				Segment 3				Segment 4			
	Nov.	Int.	Exp.	Total												
1. Patient background	32	14	4	50	8	2	6	16	2	3	3	8	4	2	1	7
2. Quantity and nature of material	21	71	77	169	2	2	5	9	2	0	0	2	1	2	0	3
3. Spatial orientation of tissue pieces	8	5	15	28	5	0	3	8	7	3	2	12	6	0	2	8
4. Artefact in image	10	9	8	27	7	9	15	31	5	1	6	12	3	6	4	13
5. Artefact in tissue	1	14	12	27	2	8	11	21	3	13	5	21	5	9	8	22
6. Findings—colour and shape	55	22	7	84	51	46	26	123	62	26	21	109	41	12	7	60
7. Findings—impression	20	5	7	32	32	1	6	39	22	7	1	30	29	8	3	40
8. Findings—name	97	46	35	178	136	111	89	336	100	96	80	276	94	68	55	217
9. Findings—architecture	11	19	14	44	6	23	21	50	17	12	17	46	16	10	8	34
10. Comparison between cases	22	0	0	22	17	0	0	17	6	0	0	6	10	0	0	10
11. Comparison within case	8	2	5	15	9	0	2	11	18	0	6	24	13	0	5	18
12. Repetition within case	2	1	4	7	5	2	8	15	13	3	7	23	13	11	21	45
13. Comparison with normal tissue	12	45	24	81	23	39	36	98	18	39	26	83	11	22	18	51
14. Absent features	7	8	10	25	13	26	19	58	11	33	21	65	13	36	27	76
15. Hypothesis	5	2	3	10	16	9	11	36	22	15	24	61	28	25	17	70
16. Diagnosis	0	1	1	2	0	4	2	6	1	9	12	22	32	56	61	149
17. Exclusion of diagnosis	0	0	0	0	0	0	0	0	1	0	0	1	6	14	3	23
18. Rejection of hypothesis	2	0	2	4	1	7	5	13	4	6	11	21	3	10	14	27
19. Heuristic comment	39	27	34	100	42	22	26	90	33	21	23	77	44	29	44	117
20. Search goal	14	3	10	27	8	8	8	24	8	9	9	26	4	1	5	10
21. Reflection—knowledge and skills	16	2	0	18	31	4	4	39	32	4	5	41	53	7	7	67
22. Reflection—diagnosis	8	4	2	14	10	4	6	20	9	6	10	25	16	8	14	38
23. Reflection—normal process	0	0	2	2	0	0	3	3	0	1	0	1	0	4	5	9
24. Reflection—solubility of slide	5	9	4	18	5	12	10	27	7	10	22	39	5	18	20	43
25. Reflection—experiment	2	2	4	8	0	1	8	9	0	0	5	5	2	0	1	3
26. Knowledge statements and inferences	10	5	8	23	32	14	17	63	28	13	11	52	32	16	17	65

intermediates in Segment 1. Knowledge statements and inferences (26) were mainly uttered by novices, throughout the whole diagnostic process.

## DISCUSSION

The aim of this study was to gain insight into the developmental stages of visual problem-solving expertise of clinical pathologists. We therefore studied the visual problem-solving process of clinical pathologists with different levels of expertise. In the Introduction, we argued that an existing model of the diagnostic process needed to be revised to fit the visual problem-solving process of clinical pathologists. We therefore introduced general problem-solving theory. To determine how this theory is reflected in the diagnostic process of clinical pathologists, microscope navigation and concurrent think aloud data were collected and analysed using a time-grid.

With regard to the first research question, ‘which expertise-related differences can be found in the temporal arrangement of the diagnostic process of clinical pathologists, as expressed in microscope navigation and clinical reasoning?’, the results of this study revealed important differences between novices, intermediates, and experts. Novices showed hardly any temporal differences in their microscope movements: Nearly all measures were on the maximal level from the first segment onwards. Their clinical reasoning shows more contrasts between segments, roughly following the pattern of comments on patient background (Segment 1), findings and comparisons (Segments 2 and 3), and reflective comments on their own knowledge and skills (Segment 4). These results suggest that novices ‘jump’ into the diagnostic process: Activity and magnification are high from the beginning, indicating a constantly intensive diagnostic process. A novice starts to search right away. However, this search has few tangible results: findings described by their appearance, few diagnoses, and many reflective comments on their abilities and knowledge statements.

The diagnostic process of intermediates shows more temporal differences than that of novices and can be divided into three parts: Segment 1, a middle part of Segments 2 and 3, and Segment 4. The first segment shows fewer movements and lower magnifications than in the middle part of Segments 2–3. In addition, there is an increase in opposed zooming movements in Segment 4. The clinical reasoning is in line with their microscope navigation: Comments on the quantity and nature of the tissue and the patient background are followed by findings and absent features. Segment 4 contains most diagnoses, exclusions of diagnoses, and hypotheses. Intermediates reflect relatively little, but if they do, it is most often on the solvability of the slide.

Whereas intermediates showed three distinct parts of the diagnostic process, experts only showed two: the first segment and the other three. During the first segment, experts used low magnification, made few opposed zooming movements, and visited few relevant areas, compared with the other segments. From the second segment onwards, magnification, zooming activity, and the hits of relevant areas increased. In the first segment of their clinical reasoning,

experts made many comments on the quantity and the spatial orientation of the tissue. Also, artefacts were mentioned in the first part of the process. The first segment contains few mentioned findings: those are concentrated in Segments 2 and 3 and so are absent findings and hypotheses. Segment 4 includes most diagnoses and rejected hypotheses. In contrast with the intermediates, the experts tended to reflect much, primarily on the solvability of the case, what they would have done in real practice, and on the experimental procedure.

With regard to research question 2, ‘how is an orientation phase reflected in the temporal arrangement of microscope navigation and clinical reasoning, and which expertise-related differences can be identified therein?’, orientation on the problem seems to play an important role in the diagnostic process of intermediates and experts. Both groups showed a relatively inactive first segment, using low magnifications and making few opposed zooming movements. Experts also visited very few diagnostically relevant areas in this first segment. Typically, utterances on the quantity and nature of tissue are made, plus some on the architecture (experts) or patient background (intermediates). Findings, hypotheses, and let alone diagnoses are not (or hardly) mentioned in this first segment. The next segments are more active and productive in the sense of findings: magnifications increase, findings are mentioned, and hypotheses are formed. There is, however, a difference between the fourth segment of experts and that of intermediates. Intermediates show an increased number of opposed zooming movements in that segment, combined with a high maximal, but low minimal magnification. This increased ‘vertical activity’—frequent zooming in and out—could mean that intermediates tend to check whether they interpreted everything correctly and drew the right conclusions. It would not be surprising if they were less confident than experts. After all, intermediates mention many diagnoses, and also many hypotheses in the last segment.

Based on their microscope navigation, the novices do not seem to orient themselves on the problem: They immediately start searching for cues at high magnification and zoom out and back in again to continue their search. However, their focus on the patient background and previous cases in their reasoning could be interpreted as an attempt at a problem representation, with the limited means they have at their disposal. This representation is thus based primarily on circumstantial information: cases just seen and information provided by the experiment.

Finally, our results provide an answer to the third research question: ‘How do the results of this study inform a new, more encompassing model of the visual problem solving process of medical specialists?’. They indicate that the visual problem-solving process of expert and intermediate clinical pathologists corresponds with general problem solving as described by Newell and Simon (1972). These two expertise levels both show a distinct phase of orientation, in which a problem representation is constructed. According to our data, this phase consisted of roughly one fourth of the diagnostic process. The ‘cut’ between the first, orientation segment and the rest of the diagnostic process was most articulate in the group of experts. Intermediates showed

similar behaviour but did focus on the clinically relevant areas from the start. During their orientation, they thus seemed to be focused more on the relevant parts already. Novices showed no clear signs of the construction of a problem representation.

In accordance with the literature on the quality of cognitive schemata used, the findings mentioned by novices contained many superficial findings, reflecting the theory of Boshuizen and Schmidt (1992, 2008) on expertise-related differences in the knowledge structures of medical diagnosticians. Also, novices make most knowledge statements, which indicate their need to reason to come to their representation. Intermediates and experts, on the contrary, report many findings reflecting the activation (comparisons with mental models of normal tissue and hypotheses) and testing (absent features) of schemata. Intermediates seem to be focused on reporting findings and drawing conclusions from them, based on their relative infrequent reflections. The fact that experts reflect on the process and task indicated that, although they are involved in the task, they keep the overview over the whole task.

Although our data correspond with the general problem-solving theory, they do not necessarily refute the more visual model by Nodine and Kundel (1987). It could well be that holistic viewing is performed by expert diagnosticians in the very first seconds of the diagnostic process, offering them an advantage over non-expert diagnosticians. However, the orientation of expert clinical pathologists seems to involve more than just a quick scan on abnormalities: With no specific instructions—other than to act as they normally would—they often counted pieces of tissue on the slide and took their time to form a representation. Additionally, our results do not refute the more advanced skill of pattern recognition of experts, which underlies holistic viewing. On the contrary, the low magnification applied by experts in the first segment—lower than that of intermediates and novices—could mean that they need less visual detail to construct their problem representation. This is in accordance with Van Meeuwen et al. (2014), who found that expert air traffic controllers apply visual chunking more than intermediates and novices.

This study is the first to take into account temporal differences within diagnostic processes of visual medical specialists, while focusing on both visual and cognitive processes in combination. Our methods offer the possibility to perform such a study without compromising on the number of participants or trials. As a result, the visual problem-solving process of diagnosticians with different levels of expertise could be studied in depth.

### Implications

The insights into the developmental stages of visual problem-solving expertise gained in this study can help educators during their training of residents. For example, they might stimulate the resident to construct an 'objective' problem representation, without checking relevant areas. Similarly, starting residents—with a level comparable with that of the novices in this study—could be scaffolded in such a way that they adopt adequate problem-solving strategies

more quickly. Their diagnostic process could be decomposed into several steps, for example, first orientate on the nature and quantity of the tissue pieces, and how they relate to each other. Besides, this study used data on microscope movements that are recorded by all virtual microscopes. For this study, several variables were designed to quantify these movements. When performed during the diagnostic process in practice, these analyses could inform pathologists and residents on their viewing behaviour and thus serve as a quality check.

In this study, we have successfully applied a theory on general problem solving to an example of visual problem solving, clinical pathology. An important question is whether this was a unique endeavour for clinical pathology, or whether these results are applicable to other domains of visual problem solving as well. Although this question is beyond the scope of this study, we have no reason to assume that these results are unique to clinical pathology. An important condition for transfer to other domains is that the task that is under study is both visually and cognitively demanding: It needs to involve complex visualisations and an ill-structured problem. This condition would rule out tasks such as visual detection tasks.

### Limitations of this study

An important limitation of this study is the use of two different ways to segment the microscope navigation data and the clinical reasoning data. The microscope navigation data were segmented using a time-grid: The time-on-task was divided into four equal segments. However, the clinical reasoning data consisted of transcripts of concurrent thinking aloud. These written sources were not time-coded and thus could not be segmented based on time. To solve this issue, all statements in a transcripts were numbered based on the chronological order: the first statement being number one, the next number two, and so on. By dividing the total number of statements by four, a grid was created. Although this solution erases the direct link between the two data sources, this way of segmenting these data still allows one to study phases in the clinical reasoning. Moreover, participants were encouraged to keep their thinking aloud constant throughout the diagnostic process. Therefore, there is an association between the time that is passed and the number of statements uttered. Therefore, we believe it still is a valid method to use.

This study has argued for the inclusion of problem-solving theory into the model of the diagnostic process of clinical pathologists. Including this theory has helped to understand the developmental stages of visual problem-solving expertise of clinical pathologists. These insights can help to improve the training of newcomers on the job by stimulating them to adopt expert visual problem-solving strategies.

### REFERENCES

- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16(2), 153–184.
- Boshuizen, H. P. A., & Schmidt, H. G. (2008). The development of clinical reasoning expertise. In J. Higgs, M. A. Jones, S. Loftus & N. Christensen

- (Eds.), *Clinical reasoning in the health professions* (pp. 113–121). Philadelphia, PA: Butterworth-Heinemann.
- Chi, M. T. H., & Glaser, R. (1985). Problem-solving ability. In R. J. Sternberg (Ed.), *Human abilities: An information processing approach*. New York: W. H. Freeman.
- Chi, M. T. H., Glaser, R., & Rees, E. (1981). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Crowley, R. S., Naus, G. J., Stewart, J., & Friedman, C. P. (2003). Development of visual diagnostic expertise in pathology—An information-processing study. *Journal of the American Medical Informatics Association*, *10*(1), 39–51.
- Gifi, A. (1990). *Nonlinear multivariate analysis*. Chichester, UK: Wiley.
- Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser & M. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum.
- Mello-Thoms, C., Hardesty, L., Sumkin, J., Ganott, M., Hakim, C., Britton, C., ... Maitz G. (2005). Effects of lesion conspicuity on visual search in mammogram reading. *Academic Radiology*, *12*(7), 830–840.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Nodine, C. F., & Kundel, H. L. (1987). Using eye movements to study visual search and to improve tumor detection. *Radiographics*, *7*(6), 1241–1250.
- Nodine, C. F., & Mello-Thoms, C. (2010). The role of expertise in radiologic image perception. In E. Samei & E. A. Krupinski (Eds.), *Medical image perception and techniques* (pp. 139–156). New York: Cambridge University Press.
- Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. Liversedge, I. Gilchrist & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 523–550). New York: Oxford University Press.
- Shute, V., & Towle, B. (2003). Adaptive e-learning. *Educational Psychologist*, *38*(2), 105–114.
- Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*(2), 205–221. DOI:10.1002/acp.1112.
- Van Meeuwen, L. W., Jarodzka, H., Brand-Gruwel, S., Kirschner, P. A., de Bock, J. J. P. R., & van Merriënboer, J. J. G. (2014). Identification of effective visual problem solving strategies in a complex visual domain. *Learning and Instruction*, *32*(0), 10–21.