VERSITA

HUMAN AFFAIRS *24*, 151–156, 2014
DOI: 10.2478/s13374-014-0214-0

# HOW CAN HUMAN PERFORMANCE IN REAL-WORLD TEAMS BE MEASURED?[1]

HANA HARENČÁROVÁ

*"A major contribution of the naturalistic decision making (NDM) community has been to describe how people actually make decisions in real-world settings."*

(Gary Klein, 2008, p. 456)

As the citation suggests, with naturalistic decision making (NDM) research has shifted from the laboratory to real-world settings, specifically to a world with changing goals, time pressures, uncertainty, unclear or insufficient information, high stakes and a dynamic environment. A new era of decision-making research began following the discovery in 1986 that firefighters do not make decisions after weighing up all the possible options, but by identifying the most suitable option.

Researchers working in decision-making are mainly interested in expert performance, or novice-expert comparisons. In these areas there is often little tolerance of mistakes (air traffic controllers, military commanders, pilots, surgeons, nurses, and so forth). Therefore, these professions demand good training and excellent performance. To conduct effective training, we require methods to assess performance—training outcomes. As some of these professions involve teamwork, we need to asses complex team interactions. The notion of macrocognition is used in NDM research for this purpose. While core cognitive research focuses on memory, attention, and so forth, that is, microcognition, macrocognition attempts to capture complex cognition while dealing with real life problems. Schraagen, Klein and Hoffman defined macrocognition "*as the study of cognitive adaptations to complexity*" (italics in original, 2008, p. 9). It is comprised of six functions (naturalistic decision making, sense-making, planning, adaptation, problem detection, and coordination) and six supporting processes (maintaining common ground, developing mental models, mental simulation and

[1] Patterson, E.S. & Miller, J. E. (Eds.). (2010). *Macrocognition metrics and scenarios.* Surrey and Burlington: Ashgate, xxxi + 307 pp.

story-building, managing uncertainty and risk, identifying leverage points, and managing attention). These are most often measurable on both individual and team levels. As the book under review here promises, the reader might expect complex methodology to measure macrocognitive function and processes. But that is not the case, the team of authors comprises 39 specialists, mostly from the domains of naturalistic decision-making, human factors, cognitive engineering, and so on, which means the book contains highly relevant knowledge from recent research in the field. However, this publication is aimed more at researchers and practitioners with specific questions than novices.

The book aims to tackle the questions of what to measure and how to measure it while assessing macrocognition as a reaction to the complexity of the organizational environment in modern companies. The goal of macrocognitive research is to understand the cognitive work of individuals or teams and to create tools to facilitate it by providing accurate metrics and measures.

The importance of measures is found in their impact on the behavior of participants. By setting measures we also set priorities, we communicate what is important, what should be focused on, and where participants should direct their efforts. The danger of inaccurate measures is that we may change something, but we may be not able to define what we have changed and whether it was a desired outcome.

In the opening chapter, *Theory Concepts Measures but Policies Metrics*, Robert Hoffman argues that we often measure things that are easily measurable but they are not necessarily the most important things to measure. Therefore it is important to develop measures of what is really relevant in performance. Hoffman argues that metrics are a matter of policy, while measures are theory-laden. This distinction enables us to concentrate on the difference between these two while creating them.

The three parts of the book, *Theoretical foundations, Macrocognition measures for real-world teams* and *Scenario-based evaluation approaches*, lead us from the basic concept of macrocognitive measures through the application of different methods before finally focusing on frequently used scenario-based training tools.

Gary Klein begins part two, *Macrocognition measures for real-world teams*, by summarizing macrocognitive functions and processes, where the first support the latter. By also providing an overview of the research paradigms for functions he enables the reader to better understand the following list of macrocognitive measures.

Before introducing the measures, we should motivate the participants to spend time completing them. In this context, Burke et al. refer to face validity as the degree to which the measure appears to assess the content. They argue that individuals need to see the relevance of the measure and how it relates to the construct of interest and their willingness to dedicate time and effort.

We will begin with a non-professional example—driving skills. Most people in industrialized societies drive a car. Research indicates that most people consider themselves to be better drivers than others (statistically impossible). This belief is often attributed to a superiority or overconfidence bias. Gerd Gigerenzer (2008) offers a different perspective on this—ecological rationality. Whatever the cause, mortality on the roads is still relatively high. In his chapter *Demand calibration in multitask environments: Interactions of micro and macrocognition* Lee argues that drivers have few opportunities to gain feedback and

therefore they may not be aware of their potentially dangerous driving practices for many years, that is, until they crash. Without feedback, there is a little chance of them learning from experience. He assumes that crashes while driving caused by distraction are an example of the failure to balance the cognitive demands associated with the road and those with in-vehicle information systems (such as, GPS, MP3 players). The cognitive demands are considered to be safety-critical, while activities related to in-vehicle information systems are regarded as competing activities. Although Lee insists on the importance of combining micro and macrocognitive perspectives on the distraction-while-driving problem, in the chapter we cannot really find the answer—the recommended measure, which would combine the macro and microcognitive perspectives.

Studying an individual's cognition is tricky, because most work is done "inside the head". Where teams are concerned, the team members have to communicate together. This allows the researchers to observe and analyze the cognition manifested in the discussion. In their chapter, *The pragmatics of communication-based methods for measuring macrocognition*, Cooke and Gorman think that "much of team cognition can be understood by studying team communication or, more generally, team interactions" (p. 161). The question often arises as to what materials should be included in data collection and what should be included in the analysis. The authors have identified the four most useful types of communication data for analysis: audio, chat, email, and logged communication events. After reviewing the common pitfalls of communication data analysis, they provide guidelines for effective analysis of team communication. One piece of advice they give is to use a Wedding cake strategy, consisting of three steps. The first step is to select interesting data using the cheapest and fastest methods. Secondly, the interesting data should be processed using moderately costly methods and finally, the researcher should run a detailed (and expensive) analysis on the selected data.

One method for capturing team interactions is observation. It would have been beneficial if Künzle et al. had mentioned how the research studies were selected in their review of observational methods in health care teams entitled *Survey of healthcare teamwork rating tools: Reliability, validity, ease of use, and diagnostic efficacy.* They provide an exhaustive assessment of validity, reliability, ease of use, and diagnostic efficacy. As they note, few authors provide information on the latter two.

Another example from healthcare is *Measurements Approaches for Transfers of Work During Handoffs* by Patterson and Wears, who have identified the transfer of work during handoffs as a perfect opportunity for researchers to gain insight into difficult areas of healthcare. The chapter includes a literature review of patient-handoff quality measurements and the authors have also created four groups of handoff quality measures: outcomes, content of interactions, interaction processes and learning. The six conceptual framings for patient handoffs are defined as Information processing, Stereotypical narratives, Resilience, Accountability, Social interaction and Cultural norms. These framings are intended "to point out different types of quality improvement interactions" (p. 138).

When we want to assess a novice's learning we can easily compare his or her results to those of an expert. But how can we compare experts? Are they at the same level? Does the performance of a particular expert differ over time? Does he become even better? Many performance measures are based on error rates. However, error rates can be tricky in measuring expert performance, where the error rate is very low, or even zero. This does not necessarily

mean that there are no differences between experts or that they do not improve their performance as they gain more experience. Shanteau et al. in their study *Assessing Expertise When Performance Exceeds Perfection* suggest resolving this problem by using the Cochran-Weiss-Shanteau index or CWS index as it is referred to (originally from Weiss & Shanteau, 2003). This index combines the measure of discrimination and consistency in judgments and is effective in revealing the differences between experts or between their performances over time. The authors provide evidence of performance amelioration in a longitudinal study of student air traffic controllers. When comparing the CWS index with other types of measures, the former provide more information than the latter criteria. These results suggest that raw measures of time or errors can be insufficient in revealing improvements in expert performance. On the other hand, the authors admit that when experts develop a new strategy, the CWS score may be lower, because of inconsistencies in their behavior or responses.

The last example of measuring macrocognition reviewed comes from the domain of analytical rigor. In their *Measuring Attributes of Rigor in Information Analysis*, Zelik, Patterson and Woods attempt to answer the question of "how to codify the rigor of an information analysis process from a macrocognitive perspective?" (p. 75) by proposing eight attributes of analytical rigor and providing descriptions and three possible scores (low, moderate, high). Their approach has the potential to reduce the risk of a shallow analysis and can also be used as a macrocognitive measure of analytical sense-making activity.

From a more conceptual point of view, Fiore et al. distinguish between team cognition and macrocognition as follows: "...team cognition tends to focus on coordinating actions between individuals, for example, understanding how team members are able to sequence their actions in service of their team tasks. But macrocognition in teams focuses more on the *knowledge work* done by a team" (italics in original, p. 179). Their article, *From Data, to Information, to Knowledge: Measuring Knowledge Building in the Context of Collaborative Cognition*, tackles the process of knowledge building. They define the context (receiving data, grounding to obtain information and forming a whole—the context) and integration (comprehending the data, organizing it as information and integrating it into knowledge) as crucial components of the problem solving process. They distinguish data (facts presented without context) from information (organized or structured data related to the context). Knowledge is created through the integration of information related to the context—creating something new. For information to become knowledge it must be integrated in such a way that it is actionable.

The next concept presented in macrocognitive measures is intent. Shattuck highlights the role of intent in *Assessment of Intent in Macrocognitive Systems*. Explicit intent comprises five successive components (formulation, communication, verification, interpretation and implementation). However, no clear link between intent and macrocognition is established. Moreover, macrocognition is referred to as a whole; it is not broken down. The author discusses the measurements according to the five steps and provides examples from the military.

Research on macrocognition features strongly in application. How can teams be trained in naturalistic settings? How can we help teams make the journey from novice to expert? How can we develop valid and reliable training tools?

These questions are partly answered in the third part of the book. Roth and Eggleston (*Forging New Evaluation Paradigms: Beyond Statistical Generalization*) question the use

of different tools designed to help practitioners to act and decide in naturalistic settings. They claim that the expected improvements are not necessarily always the outcome. The problem lies in the unsatisfactory evaluation of potential improvements. According to the authors, traditional approaches to evaluating programs in the experimental paradigm are a trade-off between internal and external validity. Their solution is to extend the external or ecological validity of the tool through work-centered evaluation. The authors argue that in joint cognitive system evaluation it is not always possible to meet requirements for statistical power (for example large samples), which reduces the significance. Studies in naturalistic decision-making are often conducted using real world teams (such as power plant operators, military staff, air traffic controllers or nurses and other medical teams). This sometimes implies that there will be a potentially low number of participants in the scenarios and also that crew-testing and scenario-running may be time-consuming (hours or days). "Work-centered evaluations emphasize the use of test cases that are representative of the range of tasks, cognitive activities, work contexts, and sources of complexities that are likely to arise in the domain and ecologically valid performance measures" (p. 211).

So how can we develop a good scenario measure? Patterson et al. provide a thorough list of the factors that increase the complexity of the scenario. These factors are clearly linked to macrocognitive functions and each is accompanied by a description and example. For this reason the chapter entitled *Facets of Complexity in Situated Work* is very useful for all those willing to create or adapt their own scenario-based tools.

Altogether, this volume, which is the sequel to *Naturalistic Decision Making and Macrocognition* (Schraagen et al., 2008), contains valuable knowledge from the area of macrocognitive measurements. Much of the book refers to macrocognition as a whole, which may cause confusion when defining which macrocognitive processes or functions are actually measured. It is frequently argued that macrocognitive processes and functions are interrelated and that it is not possible to separate them strictly. Nonetheless, some authors have proved that this distinction is possible. The volume was not intended to deal with the measures and metrics of macrocognitive functions and this is an area that remains to be covered.

As NDM research is conducted in different work environments, the methods and metrics often have to be adapted, changed or created. We now have a rich source of inspiration and advice for that very purpose. So the challenge now is not to become lost in the variety of possible methods, theoretical approaches and paradigms, but to be aware of our goals and to select or create the right one for our purposes. This will enable practitioners to improve their performances and enjoy good cognitive work, which is, according to the pleasure principle, a system which "instills in the human a sense of joyful engagement" (p. 18).[2]

## References

Gigerenzer, G. (2008). *Rationality for mortals: How people cope with uncertainty*. New York: Oxford University Press.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456-460.

[2] This article is the outcome of a project entitled "Decision making of professionals: Processual, personality and social aspects," APVV-0361-12.

Schraagen, J. M., Klein, G., & Hoffman, R. R. (2008). The macrocognition framework of naturalistic decision making. In J. M. Schraagen, L. G. Militello, T. Ormerod, & R. Lipshitz (Eds.). *Naturalistic decision making and macrocognition* (pp. 3-25). Hampshire: Ashgate.

Schraagen, J. M., Militello, L. G., Ormerod, T., & Lipshitz, R. (Eds.). (2008). *Naturalistic decision making and macrocognition*. Hampshire: Ashgate.

Weiss, D. J., & Shanteau, J. (2003). Empirical assessment of expertise. *Human Factors*, *45*(1), 1-13.

Institute of Experimental Psychology,
Slovak Academy of Sciences,
Dúbravská cesta 9,
814 04 Bratislava,
Slovakia
E-mail: h.harencarova@gmail.com