Applied Ergonomics xxx (2016) 1–8

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Rasmussen and the boundaries of empirical evaluation

Penelope Sanderson [a, *], Catherine Burns [b]

[a] The University of Queensland, Brisbane, Australia
[b] University of Waterloo, Waterloo, Canada

ARTICLE INFO

Article history:
Received 3 October 2016
Accepted 6 October 2016
Available online xxx

Keywords:
Jens Rasmussen
Empirical evaluation
Constraints
Work analysis

ABSTRACT

In this special issue, many of the papers focus on Rasmussen's analytic contributions to the understanding of work in complex sociotechnical systems. Work is analysed for the purpose of developing new designs that can improve the nature of that work. The evaluation of such designs was a key part of Rasmussen's program, yet he was often sceptical of the claims made for the generalizability of empirical studies. To tackle this problem, he extended his work analysis framework to provide a way of thinking about empirical evaluation. As authors of this paper, we come from two different backgrounds—systems engineering in the case of Burns, and engineering psychology in the case of Sanderson—and over the decades of our respective research programs, we have both performed many empirical investigations: field investigations, simulation studies, and behavioural laboratory experiments. Rasmussen's scepticism—and his writings on the issue—have stimulated and shaped our own research. In this brief paper we present our interpretation of Rasmussen's perspective, we provide examples how our research sits within Rasmussen's framework of constraints defining boundary conditions for experiments, and we draw conclusions for the future.

> "… it is fashionable to make distinctions between the laboratory and the world but it is apparently not so fashionable to make an argument for linking them." Rasmussen et al. (1994) p. 228.

The papers in this special issue celebrating the legacy of Jens Rasmussen have covered a range of topics that reflect the abiding interests, concerns, and theoretical positions of their authors. Some authors have provided us with a history and analysis of key themes from Rasmussen's work—examples include Waterson et al.'s (2016a) extraction of eight key themes of Rasmussen's work from four key papers, Kant's survey of themes in Rasmussen's work from 1961 to 1986, and Wears' (2016) bibliometric analysis of Rasmussen's influence. Further special issue authors have related Rasmussen's work to parallel intellectual traditions, such as Le Coze's (2016) review of the cybernetic foundations of Rasmussen's work, and Flach's (2016) and Bennett's (2016) positioning of Rasmussen's work in the context of a triadic semiotic model. Leveson (2016) positioned Rasmussen's work in relation to the increasing application of general systems theory to safety in the face of the growing complexity of sociotechnical systems, especially with software-based control systems and Sheridan (2016) assessed the work in terms of requirements for successful models. Other authors have pointed to the moral consequences of reformulating system safety as a systems problem, rather than one based in the abilities and dispositions of workers—Dekker (2016) with his account of a second victim story, and Kant's (2016) review of system design being a joint responsibility of operators and designers.

The special issue has also covered critiques of aspects of Rasmussen's work, areas of misinterpretation of the work, and areas of limitation. For example, Le Coze (2016) identified different forms of constructivism that are evident in Rasmussen's work, but that also challenge it. Flach (2016) noted dangers in too-strict adherence to some of the conventional analytic templates that Rasmussen developed for Cognitive Work Analysis (CWA) and accident analysis. Difficulties in applying the phases of CWA, and especially Work Domain Analysis (WDA) were reported by Hilliard and Jamieson (2016) as they developed information systems for energy efficiency monitoring and targeting. Two Accimap analyses of the South Korea Sewol Ferry accident produced somewhat different outputs (Kee et al., 2016; Lee et al., 2016; Sharples, 2016) and Salmon et al. (2016) noted the difficulties of conveying the philosophy underlying Accimaps to potential domain end-users.

Finally, many papers in the special issue address adaptations and extensions of Rasmussen's work. For example, after an

* Corresponding author.
E-mail address: p.sanderson@uq.edu.au (P. Sanderson).

http://dx.doi.org/10.1016/j.apergo.2016.10.003
0003-6870/

extensive review of uses of Accimaps since 2000, Waterson et al. (2016a, b) noted ways that Rasmussen's original ideas have been adapted and "remixed" with other methods to meet the needs of specific case studies, and they identified the advantages and disadvantages of doing so. In a novel adaptation of the ideas of abstraction and aggregation, Rouse et al. (2016) developed interactive visualisations to support 'enterprise diagnostics' about product viability. Leveson (2016) reported the use of the AH and risk management hierarchy to document constraints in the design process during system specification. Finally, Naikar (2016) showed that the ideas underlying CWA can be extended to develop new ways to organise work across multiple actors, and can even apply beyond traditional human factors and engineering to the development of military doctrine.

1. The present paper

In the present paper we focus on an aspect of Rasmussen's work that has not been covered in depth in the special issue – his thinking about evaluation and especially about empirical evaluation. Understanding how to study complex systems interventions, and to do so with an appropriate balance of fidelity, control, and generalizability, is a challenge to many researchers in the field of human factors (Sanderson and Grundgeiger, 2015). As co-authors of this paper we have different backgrounds—Burns is a systems engineer and Sanderson is an engineering psychologist—but we each conduct many empirical investigations, some of them formally experimental in nature. In this paper we examine some of Rasmussen's thinking on empirical studies of all kinds, noting and using the formalisms and analytic templates he introduced to illustrate his thinking.

Rasmussen's view of the shortcomings of experimental psychology studies have been noted in a few places in this special issue, including Flach (2016). As the quote at the start of the paper indicates, Rasmussen recognised that there was not a systematic way of identifying and relating the contributions of investigations done under controlled laboratory conditions, and investigations performed in authentic work situations. Alongside this was a concern about a tendency for the generalizability of some empirical studies to be overstated by their authors, or not stated at all. A key motivation for Rasmussen's work in the area was his need to develop concepts that would help analysts identify how generalizable their conclusions could be about the effectiveness of new information systems or new work arrangements. Ways of including and analysing the human operator's engagement with content were just as important as understanding psychological processes, and Rasmussen frequently invoked Brunswik's (1952; 1956) ideas on representative design of experiments to argue that sampling tasks and situations should be as important as sampling participants, as in the quotation below.

> Another line of analysis is the study of mental strategies and of human subjective preferences in the specific real life tasks. This area raises the problem of experimental methods within psychology. We need reliable methods for probing cognitive control structures and mental models during real task performance and during experimental conditions. … This apparently requires a new direction within experimental psychology to include complex experiments in the laboratory repertoire, and it requires that psychologists not only focus their interest upon the human but include detailed analysis of the human's task environment. This development was foreseen by Brunswik in 1952 when he advocated equal attention by psychologists to the real life task content and to the psychological processes of the performer (Rasmussen, 1986; p. 191)

2. Rasmussen's approach to evaluation

Ultimately, and almost from the outset, the goal of much of Rasmussen's work was to specify the form that information systems should take so the human operator could more effectively execute the intent of the designers (Rasmussen, 1968a, 1968b, 1968c). In many cases, this specification was being performed for first-of-a-kind systems, raising profound problems in terms of how evaluation could be performed.

For this task, Rasmussen advocated a mixture of analytical and empirical evaluation. Analytic evaluation would provide a means of in-principle evaluation using the analytic products from the phases of CWA, as seen for example in case studies such as that by Naikar and Sanderson (2001). Empirical evaluation would provide evidence of the effectiveness of new design concepts—but it fell short. Rasmussen and his colleagues could not easily or effectively draw insights from much of the existing corpus of experimental research on human performance, diagnosis, and decision making.

Rasmussen noted that designers need more explicit descriptions of the *boundary conditions* of experiments, which would then help them transfer the findings to their work contexts in a more precise way. Much of Rasmussen's thinking found its fullest expression in the Rasmussen et al. (1994) book which was written once most of the important analysis and empirical evaluation had been done on the Bookhouse project (Pejtersen, 1989) and we draw much of the following from the former source.

Rasmussen's approach to specifying the generalizability of an empirical investigation was to be very specific about the constraints that might vs. might not be operating on an actor in the context of an empirical evaluation. The constraints in operation in an empirical study will dictate the situations and context to which the results of the evaluation can generalise. To provide tools that would help analysts specify the status of those constraints, Rasmussen turned to several of his work analysis frameworks.

We present three sets of analytic tools that Rasmussen used to describe his ideas. Because the insights from the tools are intertwined, we present them first in their abstract form, and then illustrate them with examples from our own research. Finally, we make recommendations for how aspects of Rasmussen's ideas might be used to greater effect in future research that is intended to contribute to our understanding of human-system integration.

2.1. Behaviour-shaping constraint boundaries

The first and most general framework is shown in Fig. 1 (Rasmussen et al., 1994, Figure 8.5) which shows all the task-related and agent-related factors that shape the behaviour that human actors exhibit as they work within a system. The uppermost boxes reflect the tasks that emerge from the purpose of the system, given the system's functional structure. The lower boxes reflect properties of the human actors, given their cognitive and motivational capabilities and their organisational context.

The ellipses superimposed over Fig. 1 represent the extent to which the human actors' behaviour will be constrained in different kinds of experimental studies. In each case, the factors that are constrained (assumed, or held constant) in an experimental study lie *outside* the ellipse, and the factors from which the human actor—or experiment participant in some cases—is free to choose lie *inside* the ellipse.

So, for example, the ellipse labelled "2. Constrained cognitive task" is a boundary indicating that an experiment has specified, through explicit instruction, what the participant's cognitive task will be. An experiment with such a boundary is intended to remove

Please cite this article in press as: Sanderson, P., Burns, C., Rasmussen and the boundaries of empirical evaluation, Applied Ergonomics (2016), http://dx.doi.org/10.1016/j.apergo.2016.10.003