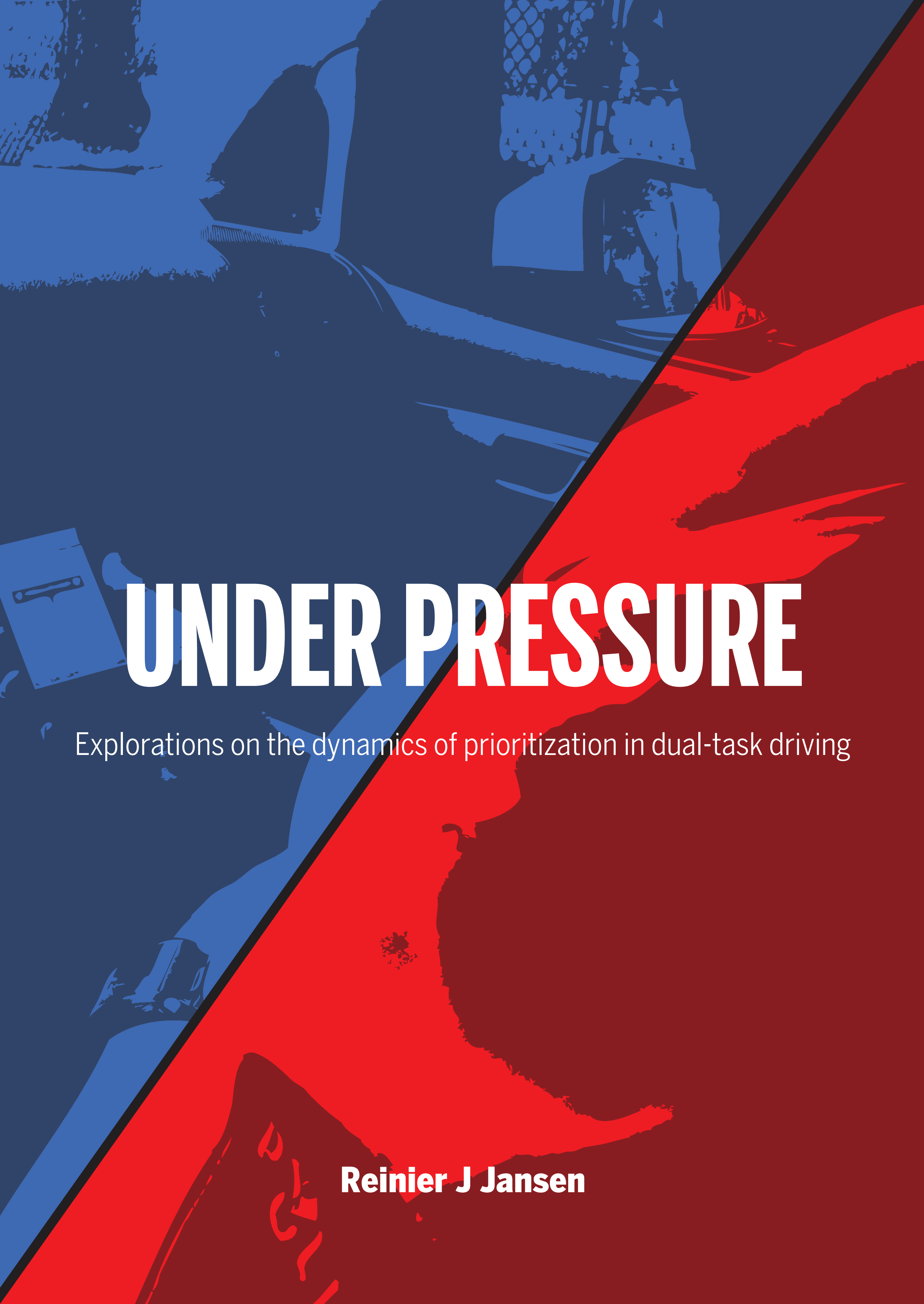

UNDER PRESSURE

Explorations on the dynamics of prioritization in dual-task driving

Reinier J Jansen

This page has intentionally been left blank.

UNDER PRESSURE

Explorations on the dynamics of prioritization in dual-task driving

Printed by: Delft Academic Press

ISBN/EAN: 97890065624086

UNDER PRESSURE

Explorations on the dynamics of prioritization in dual-task driving

Proefschrift

ter verkrijging van de graad van doctor

aan de Technische Universiteit Delft,

op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben;

voorzitter van het College voor Promoties,

in het openbaar te verdedigen op vrijdag 23 juni 2017 om 10:00 uur

door

Reinier Johan JANSEN

Ingenieur industrieel ontwerpen,

geboren te Leiden

Dit proefschrift is goedgekeurd door de

Promotor:	Prof.dr. H. de Ridder
Copromotor:	Dr. R. van Egmond

Samenstelling promotiecommissie

Rector Magnificus	Voorzitter
Prof.dr. H. de Ridder	Technische Universiteit Delft, promotor
Dr. R. van Egmond	Technische Universiteit Delft, copromotor

Onafhankelijke leden

Prof.dr. P.A. Hancock	University of Central Florida, USA
Prof.dr. N.A. Taatgen	Rijksuniversiteit Groningen
Prof.dr. M.P. Hagenzieker	Technische Universiteit Delft
Prof.dr. P.J. Stappers	Technische Universiteit Delft
Dr. N. van Nes	SWOV

This PhD research was partially funded by Staf Korpsleiding/Directie Facility Management, Nationale Politie, and partially funded by the European Space Agency (project code AO 1–7223/12/F/MOS).

to Jos & Betty

Contents

Chapter 1

Introduction

> *"Keep off of our weekends!"* (in: Van Loon & Klumpenaar, 2015)
>
> *"Look at what colleagues have been confronted with the past weeks. These are situations one can only handle well rested."* (in: Boomsma, 2011).
>
> *"Cost reductions, reorganizations, performance targets, operational pressure; too much is happening at the same time."* (in: Van Es & Stoker, 2015)

Under pressure - it is a feeling many of us experience in our daily lives. The above protests could have been taken from many professional domains, such as healthcare and education. In this case, however, the protests have been voiced by police officers, and their overwhelming multi-task environment while on patrol provides the stage for this thesis.

As with nurses and teachers, many police officers suffer from burnout (Van der Steur, 2016), caused by staff shortage, lack of adequate work material, and work overload (Kop & Eeuwema, 2001). Therefore, if hiring more staff is not an option, the work material should be improved to mitigate the effect of work overload on burnout. What demands should be made of the work material to achieve this goal?

It is often thought that superimposing layers of information through information technology will make effortful tasks easier. However, in the vision of Ambient Intelligence (Aarts & Marzano, 2003; Markopoulos, 2016), such an approach may prove to be counterproductive, unless the technology is context-aware (i.e., involving time, location, environmental factors, social interactions) and personalized (i.e., tailored to the user's needs). Streefkerk et al. (2006) argue that these principles also apply to information technology for police officers patrolling on foot. An observation study has shown that police officers forget seventy percent of the information presented during a briefing (Scholtens et al., 2013). Later that year, a dedicated smartphone application was developed for officers on the street (Schalkwijk, 2013), aiming to provide the right information at the right time (e.g., to know whether a traffic offender is also a wanted person).

Is it also appropriate to push such information to police officers when they are driving their car, instead of patrolling on foot, and if so, under which circumstances? Police work involves perpetual transitions between different types of driving activities (Sørensen & Pica, 2005), such as surveillance, pursuit, and emergency response. Within this fragmented work setting, police officers are expected to continuously inform themselves. However, such an in-vehicle task has been associated with increased accident risk (Caird et al., 2008; Dingus et al., 2016; Lee et al., 2001; Strayer et al., 2003). Therefore, context-awareness of the information technology becomes even more important in a dynamic driving context. Such a design challenge commonly occurs when developing information technology as part of a complex socio-technical system: there are numerous actors and interconnections (e.g., driver, co-driver, vehicle, information technology, organization, citizens), and their dynamic interactions and interdependence are difficult to describe, understand, predict, and change (Magee & de Weck, 2004). According to Norman (2010) good design can help tame the complexity of information technology in two ways. First, by making the underlying logics of technology understandable. Second, by connecting technology to our skills and abilities.

One such skills is the use of task prioritization, which is the process of allocating attention to one task at the expense of another task (Gopher et al., 1989). Everybody makes use of task prioritization, but some are better in meeting situational demands than others. It is with task prioritization that a pivotal difference emerges between police officers and regular drivers. Contrary to police officers, regular drivers can choose to ignore incoming messages. As a result, police officers are expected to prioritize differently between the tasks (i.e., driving, interacting with information technology) than regular drivers. Can police officers live up to this expectation? The influence of task prioritization has received limited attention in previous traffic research, possibly because the topic is less relevant for the majority of drivers. This is indicative of a widespread implicit assumption that drivers will naturally prioritize the driving task over other in-vehicle tasks. This assumption has not been addressed, until now.

The scientific goal of this thesis is to understand the mechanisms that underlie and/or result from task prioritization in a dynamic complex socio-technical system, such as the police context. From an applied perspective, the goal is to investigate the leeway to push information to Dutch police officers when they are driving their vehicle in varying work situations. A cross-disciplinary approach has been taken, in which ethnographic field studies inform a series of controlled laboratory experiments. The results of this endeavour include a novel method to describe workflow fragmentation, as well as theoretical contributions to existing models on task performance and coping behaviour.

1.1 Background operational policing[1]

The studies reported in this thesis are part of a project on the information environment of Dutch police officers. This section briefly reviews their work to outline the project scope.

The Dutch police organisation consists of one national unit, and ten regional units. The national unit deals with, e.g., highway patrol, organized crime, and terrorism. In addition, two types of police work can be found in a regional unit (Stol et al, 2004): community policing and operational policing. Community policing is pro-active and preventive, and involves considerable time on networking with civilians (Stol et al., 2004; Smith et al., 2001). The focus in this thesis, however, is on operational policing. This type of police work is mostly reactive: it is time and safety critical work based upon officers attending incident sites by car (Sørensen & Pica, 2005). When an operational police team in The Netherlands is not assigned to an active call, officers typically spend their time on criminal investigation (i.e., based on assignments handed out during briefing), or law enforcement (e.g., surveillance, traffic control). For this type of work, each police station employs a number of concurrently operating police vehicles. Two officers usually occupy one vehicle, although some regions are experimenting with additional solo patrol vehicles to cover larger areas.

The information environment of a patrolling officer consists of numerous concurrent visual and auditory channels. The patrol car is equipped with a specialized in-vehicle information and communication system. The Mobile Data Terminal (MDT) is a touch screen device positioned on the vehicle console in-between the driver and co- driver, providing the officer with a number of functionalities (see Figure 1.1). First, the vehicle is equipped with an Automatic Number Plate Recognition (ANPR) system. This system compares license plates scanned by on-board cameras with a database of delicts linked to specific number plates. In case of a 'hit', an alarm can be heard through the car's speakers, and information on the vehicle is

---

[1] This section has been adapted from Jansen et al. (2014)

Figure 1.1. View on the cockpit of a police vehicle. Interior information processing tasks: operating the Mobile Data Terminal (MDT) and mobile phones, attending incoming radio messages, communication with the co-driver. Exterior information processing tasks: monitoring the environment, including other road users (i.e., as driver, and as police officer). NOTE: even though dedicated navigation technology has been developed (i.e., MDT), officers frequently make use of their mobile phones to find their way.

displayed on the MDT. Additionally, officers use the MDT to acquire information on a person, to control the lights on top of the car, and for navigation.

Two main modes of communication between the control room and patrolling officers exist: direct contact using a mobile phone, and two-way broadcasting. Regarding the latter, officers are equipped with a portophone for radio contact, which consists of an earpiece, a microphone, and a channel selector. Additionally, the vehicle's interior loudspeakers may be used. Broadcast radio messages typically start with a numerical code consisting of the region and the team it is intended for. Consequentially, officers continuously monitor incoming codes to detect if a call is meant for them. The co-driver, if present, typically uses pen and paper to memorize details of a call, as well as observations made when dealing with a call.

Monitoring this information environment while driving may have consequences on performance, especially in case of solo patrol. Multiple

Resource Theory (MRT) predicts that time-sharing between two tasks is best when they require the use of different processing stages (e.g., cognitive vs. response), processing codes (e.g., spatial vs. verbal), and modalities (e.g., visual vs. auditory) (Wickens, 2008). However, the independence of modalities claimed by MRT has been criticized. For example, Spence and Read (2003) showed that dual-task performance decreases when the spatial location of an auditory speech shadowing task does not coincide with the spatial location of a visual driving simulator task. Since police officers typically monitor incoming messages through their earpieces (i.e., from one side), one can expect lower dual-task performance than predicted by MRT. These decrements may be enlarged when the traffic conditions become more demanding (Patten et al., 2006), for example during pursuits and high priority calls. Additionally, Anderson et al. (2005) found that police officers frequently perform more than two tasks at a given time, which may also result in performance decrements (e.g., Recarte & Nunes, 2003). Therefore, designing the cockpit of a police vehicle requires an understanding of the cognitive processes involved with in-vehicle technologies, as well as an understanding of the successive situations in which in-vehicle technologies are used.

1.2 Terminology

At the beginning of this research project the Dutch National police asked: ‘How much information can police officers process in varying work situations?’ Such a practical question reflects the assumption that people have an upper limit with regard to their capacity to process and act upon incoming information (e.g., Broadbent, 1958; Kahneman, 1973; Pashler, 1994; Wickens, 1984). This thesis frequently uses four concepts when the upper limit of information processing is (about to be) reached, namely: ‘task demand’, ‘task performance’, ‘mental workload’, and ‘effort’.

Task demand is determined by the goal that has to be attained by means of task performance (De Waard, 1996). For police officers, one of the goals in an emergency response is to reach the incident location as soon as possible.

Suppose the route involves a high traffic density, which requires many overtaking maneuvers. In that case the task demand associated with the above goal is higher than when the traffic density would have been low. Hence, task demand is viewed as an external property, independent of the person who performs the task. Task performance, then, is measured as the degree to which the task goal has been met (e.g., in the above example faster arrival corresponds with higher performance).

Mental workload is defined as the proportion of information processing capability required to perform a (combination of concurrent) task(s) (Brookhuis & De Waard, 2000; De Waard 1996, Kahneman, 1973). In this definition, mental workload is a subjective property, in that the effect of task demand on the person performing the task is mediated by, a.o., individual skills, motivation to perform a task, strategies applied in task performance, and mood (Brookhuis et al., 2009). In the emergency response example, a veteran police officer may experience the same traffic condition with a lower mental workload than a novice police officer. This does not imply that the novice police officer is unable to reach the same level of performance as the veteran police officer. On the contrary, when faced with suboptimal performance, the novice police officer may invest more effort, which Hockey (2011) views as an optional response to the perception and appraisal of task demands. Therefore, a challenge for the design of in-vehicle information technology is to ensure that investing more effort indeed remains an optional response.

1.3 Research aim

One can expect to find high levels of mental workload and effort in the context of operational policing, because work overload has been reported as one of the main organizational stressors in police work (Kop & Euwema, 2001). When a police officer attempts to perform multiple tasks, which combined approach the limits of the officer's information processing capacity, then performance on one or more tasks will suffer. That is, unless the police officer invests more effort. Consequently, tradeoffs are expected between performance, mental workload, and effort (De Waard, 1996; Young et al., 2015). This thesis investigates the influence of task prioritization on such tradeoffs.

The use of information technology has been observed in a naturalistic multi-task setting. The importance of task prioritization follows from this use. Next, mechanisms underlying and/or resulting from task prioritization have been studied in a controlled laboratory setting. Special attention has been paid to the dynamic context of police work. It should be noted, though, that work overload as a consequence of concurrent multi-tasking is not unique to the context of operational policing. For example, the phenomenon has also been reported in the domains of healthcare (Rauhala et al., 2007), teaching (Hagen, 2017), air traffic control (Brookings et al., 1996), aviation (Haeusler et al., 2012), and process monitoring (Yang et al., 2012). Therefore, the aim of this thesis is to produce knowledge on task prioritization that can be generalized across such diverse socio-technical systems.

1.4 Thesis outline

Two observation studies and five experimental studies are presented. Figure 1.2 summarizes the relation between the observation studies, experimental studies, and the chapters in this thesis.

The goal of the first observation study was familiarization with police work, with a focus on how officers interact with information systems under various work conditions. Dutch police officers were joined during their shifts. An ethnographic approach has been followed to generate ecologically valid insights on workflow dynamics. This study has resulted in the Transitional Journey Map (TJM), a novel method to visualize and quantify workflow fragmentation.

To produce generalizable knowledge, a second observation study examines whether the TJM method can be applied across different domains. The context of satellite control rooms was chosen based on two similarities with

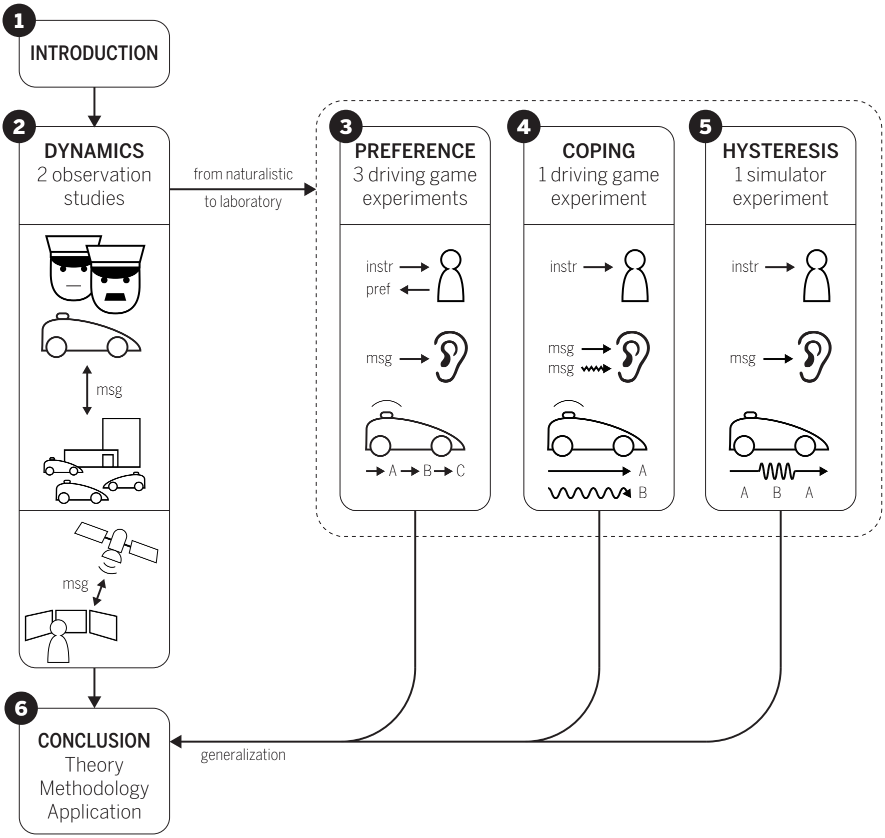

Figure 1.2. Thesis outline. The context of operational policing (e.g., multiple users, multiple tasks) has been abstracted into a dual-task context with one user. Chapter numbers are presented in black circles. NOTE: instr = priority instruction, pref = preference, msg = message.

the police context. First, both police officers and control room operators continuously monitor incoming messages. Second, both contexts feature episodes of high concurrent task demands. The observation studies are described in Chapter 2.

The goal of the experimental studies was to understand the cognitive processes underlying a particularly difficult human-machine interaction encountered in the observation study on police work: monitoring incoming radio messages while driving. The experiments followed an information-processing approach (Proctor & Vu, 2009). This approach characterizes the human as a communication system consisting of several distinct processes that operate on representations of information, mediating between perception and action. Although these cognitive processes cannot be observed directly, an understanding about them can be gained by measuring task performance, effort, and mental workload. Mental workload has traditionally been measured through three categories of parameters: measures of task performance, physiological metrics, and subjective reports (Da Silva, 2014; De Waard, 1996; O'Donnell & Eggemeier, 1986; Young et al., 2015). Regarding the first category, performance on a secondary task (e.g., phone conversations in traffic research) has been associated with spare capacity unused by the primary task (e.g., driving), but only for situations in which one task takes priority over the other (O'Donnell & Eggemeier, 1986; Young et al., 2015). In the experiments reported in this thesis, however, task prioritization has been manipulated. Therefore, task performance measures have been included, but they have not been used to infer mental workload. Instead, two common self-report scales have been used, namely the NASA Task Load indeX (NASA-TLX; Hart and Staveland, 1988) and Zijlstra's (1993) Rating Scale Mental Effort (RSME). Both scales have been proven to be sensitive to variations in mental workload in situations with high task demands (De Waard, 1996; Hill et al., 1996; Verweij & Veltman, 1996).

The scope of the experimental studies is on dual-tasking (i.e., as opposed to multi-tasking) to limit the methodological complexity. Furthermore, a focus on solo patrol has been chosen to ensure dual-tasking takes place (i.e., the co-driver in a dual patrol situation is not involved in controlling the car). Each of the experimental studies have addressed a distinct characteristic of police work by manipulating the properties of a driving task and an auditory memory task.

Chapter 3 reports a series of three experiments on the influence of prioritization preferences on following priority instructions. A driving task has been constructed using a driving game, with the goal to reach as many destinations as possible (i.e., representing police emergency response). Parallel to the driving task, a series of auditory news items have been presented, which had to be memorized (i.e., representing police radio communication). Finally, two priority instructions have been used: either to prioritize the driving task, or to prioritize both tasks equally. The results show that people differ in their preferences regarding task prioritization. These preferences can be overruled by priority instructions, but only after increased dual-task exposure. The findings have yielded a proposal for a new theoretical model to explain dual-tasking and mental effort based on the existing models of Threaded Cognition Theory (Salvucci & Taatgen, 2008) and Hockey's (1997, 2011) Compensatory Control Model.

Chapter 4 examines how priority instructions influence people's ability to cope with varying task demands. A set of methodological requirements is postulated to infer Hockey's (1997, 2011) coping strategies from tradeoffs between task performance and mental effort, taking into account the role of task prioritization. The experiment reported in Chapter 4 builds on the previous experiments by using the same priority instruction set. The tasks reported in Chapter 3 have been adapted to include multiple task demand levels, where they previously featured relatively constant task demands. The driving task featured an easy straight route and a difficult curvy route. With the memory task, the signal-to-noise ratio of the auditory news items was manipulated to mimic the often suboptimal conditions of police radio communication (e.g., distortion in portophones). Knowledge on task prioritization is shown to be essential to infer coping strategies. Furthermore, empirical evidence is presented for the existence of two coping strategies that were not previously described by the Compensatory Control Model (Hockey, 1997, 2011).

The experimental conditions in Chapter 4 were interleaved by short breaks. In reality, however, police officers frequently face transitions between different task demand levels without the opportunity of a break. For example, the observation study in Chapter 2 shows that at one moment officers may be surveilling a quiet neighbourhood, and at the next moment they may be rushing towards an emergency. Will police officers have enough time to recover from a demand transition, before they receive the next message from their portophone? And is this recovery period influenced by task prioritization?

Chapter 5 investigates the ongoing influence of demand levels prior to a sudden demand transition, a carry-over effect also known as 'hysteresis' (e.g., Morgan & Hancock, 2011). Three experimental conditions with low, high, and low task demands have been constructed by manipulating the frequency of lane changing in a driving simulator task. In addition, the auditory memory task and the priority instruction set as described in Chapter 3 have been used. Compared to previous studies on hysteresis, a novel approach is that subjective mental workload has been measured not only at the end of each experimental condition, but also during the experimental conditions. This periodic assessment has proved to be essential to understand the temporal development of hysteresis in mental workload.

Finally, Chapter 6 synthesizes the results of the observation and experimental studies in terms of theoretical and methodological contributions, followed by suggestions for future research. Furthermore, practical implications are suggested, in particular for the development of information technology in police vehicles. As such, Chapter 6 provides several important takeaway messages for fundamental researchers, as well as for designers.

References

Aarts, E.H.L. & Marzano, S. (2003). The new everyday: views on ambient intelligence. Rotterdam, 010 publishers.

Anderson, G.S., Courtney, A., Plecas, D. & Chamberlin, C. (2005). "Multi-tasking behaviors of general duty police officers." Police Practice and Research **6**(1): 39-48.

Boomsma, K. (2011). Training agent schiet tekort [Insufficient training for police officers]. Metro Nieuws**:** 4, 26 October 2011, http://www.metronieuws.nl/nieuws/2011/10/walter-welting-training-agent-schiet-tekort.

Broadbent, D.E. (1958). Perception and communication. London, Pergamon.

Brookhuis, K.A., Van Driel, C.J.G., Hof, T., Van Arem, B. & Hoedemaeker, M. (2009). "Driving with a congestion assistant; mental workload and acceptance." Applied Ergonomics **40**(6): 1019-1025.

Brookhuis, K. & De Waard, D. (1993). “Assessment of driver's workload: performance and subjective and physiological indexes.” In: Hancock, P.A. & Desmond, P. (Eds.), Stress, Workload, and Fatigue. Londen, Lawrence Erlbaum Associates, Inc.**:** 321-333.

Brookings, J.B., Wilson, G.F. & Swain, C.R. (1996). "Psychophysiological responses to changes in workload during simulated air traffic control." Biological Psychology **42**(3): 361-377.

Caird, J.K., Willness, C.R., Steel, P. & Scialfa, C. (2008). "A meta-analysis of the effects of cell phones on driver performance." Accident Analysis & Prevention **40**(4): 1282-1293.

Da Silva, F.P. (2014). "Mental Workload, Task Demand and Driving Performance: What Relation?" Procedia - Social and Behavioral Sciences **162**: 310-319.

De Waard, D. (1996). The measurement of drivers' mental workload. Traffic Research Center VSC. Groningen, University of Groningen. PhD thesis.

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M. & Hankey, J. (2016). "Driver crash risk factors and prevalence evaluation using naturalistic driving data." PNAS **113**(10): 2636-2641.

Gopher, D., Weil, M. & Siegel, D. (1989). "Practice under changing priorities: An approach to the training of complex skills." Acta Psychologica **71**(1-3): 147-177.

Haeusler, R., Hermann, E., Bienefeld, N. & Semmer, N. (2012). “How cockpit crews successfully cope with high task demands.” In: De

Voogt, A. & D'Oliveira, T.C. (Eds.), Mechanisms in the chain of safety - Research and operational experiences in aviation psychology. Ashgate**:** 47-66.

Hagen, K. (2017). “Laag salaris en hoge werkdruk jagen onderwijspersoneel weg.” [Teaching staff quit their jobs due to low salary and high work pressure] Onderwijsblad **2**: 7. Utrecht, Netherlands: Algemene Onderwijsbond.

Hart, S. G., & Staveland, L. E. (1988). “Development of NASA- TLX (Task Load Index): Results of empirical and theoretical research.” In: Hancock, P.A. & Meshkati, N. (Eds.), Human mental workload (pp. 139–184). Amsterdam, Netherlands: North Holland Press.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklade, A.L. & Christ, R.E. (1992). "Comparison of Four Subjective Workload Rating Scales." Human Factors **34**(4): 429-439.

Hockey, G.R.J. (1997). "Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework." Biological Psychology **45**(1-3): 73-93.

Hockey, G.R.J. (2011). “A motivational control theory of cognitive fatigue.” In: Ackerman, P.L. (Ed.), Cognitive fatigue: multidisciplinary perspectives on current research and future applications. Washington, DC, American Psychological Association**:** 167-188.

Jansen, R. J., van Egmond, R., de Ridder, H., & Silvester, S. (2014). “Transitional journey maps: Capturing the dynamics of operational policing.” In: D. De Waard, K. Brookhuis, R. Wiczorek, F. Di Nocera, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.), Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference (pp. 15–27). Santa Monica, CA: Human Factors and Ergonomics Society.

Kahneman, D. (1973). Attention and effort. New Jersey, U.S.A., Prentice Hall.

Kop, N. & Euwema, M.C. (2001). "Occupational Stress and the Use of Force by Dutch Police Officers." Criminal Justice and Behavior **28**(5): 631-652.

Lee, J.D., Caven, B., Haake, S. & Brown, T.L. (2001). "Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway." Human Factors **43**(4): 631-640.

Markopoulos, P. (2016). "Ambient Intelligence: Vision, research, and life." Journal of Ambient Intelligence and Smart Environments **8**(5): 491-499.

Morgan, J.F. & Hancock, P.A. (2011). "The effect of prior task loading on mental workload: An example of hysteresis in driving." Human Factors **53**(1): 75-86.

Norman, D.A. (2010). Living with complexity. Cambridge, MA, MIT Press.

Magee, C.L. & De Weck, O.L. (2004). “Complex system classification.” Fourteenth Annual International Symposium of the International Council On Systems Engineering (INCOSE). Las Vegas, NV**:** 1-18.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In: Boff, K.R., Kaufman, L. & Thomas, J.P. (Eds.), Cognitive Processes and Performance. John Wiley and Sons.

Pashler, H. (1994). "Dual-Task Interference in Simple Tasks: Data and Theory." Psychological Bulletin **116**(2): 220-244.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L. & Svenson, O. (2006). "Driver experience and cognitive workload in different traffic environments." Accident Analysis & Prevention **38**(5): 887-894.

Proctor, R.W. & Vu, K.-P.L. (2009). "Cumulative Knowledge and Progress in Human Factors." Annual Review of Psychology **61**: 623-651.

Rauhala, A., Kivimäki, M., Fagerström, L., Elovainio, M., Virtanen, M., Vahtera, J., Rainio, A.-K., Ojaniemi, K. & Kinnunen, J. (2007). "What degree of work overload is likely to cause increased sickness absenteeism among nurses? Evidence from the RAFAELA patient classification system." Journal of Advanced Nursing **57**(3): 286-295.

Recarte, M.A. & Nunes, L.M. (2003). "Mental Workload While Driving: Effects on Visual Search, Discrimination, and Decision Making." Journal Experimental Psychology: Applied **9**(2): 119-137.

Salvucci, D.D. & Taatgen, N.A. (2008). "Threaded cognition: an integrated theory of concurrent multitasking." Psychological Review **115**(1): 101-130.

Schalkwijk, L. (2013). Smartphone politie weet alles van ons [Police smartphone knows everything about us]. Algemeen Dagblad**:** 19, 28 November 2013.

Scholtens, A., Groenendaal, J. & Helsloot, I. (2013). “De operationele politiebriefing onderzocht” [Investigation of briefing operational policing]. Politie & Wetenschap. Apeldoorn, The Netherlands.

Smith, B.W., Novak, K.J. & Frank, J. (2001). "Community Policing and the Work Routines of Street-Level Officers." Criminal Justice Review **26**(1): 17-37.

Sørensen, C. & Pica, D. (2005). "Tales from the police: Rhythms of interaction with mobile technologies." Information and Organization **15**(2): 125-149.

Spence, C. & Read, L. (2003). "Speech Shadowing While Driving: On the Difficulty of Splitting Attention Between Eye and Ear." Psychological Science **14**(3): 251-256.

Stol, W.P., van Wijk, P., Vogel, G., Foederer, B. & van Heel, L. (2004). Politiestraatwerk in Nederland: Noodhulp en Gebiedswerk [Police streetwork in The Netherlands]. Apeldoorn / Zeist, Nederlandse Politie Academie / Kerckebosch.

Strayer, D.L., Drews, F.A. & Johnston, W.A. (2003). "Cell Phone-Induced Failures of Visual Attention During Simulated Driving." Journal of Experimental Psychology: Applied **9**(1): 23-32.

Streefkerk, J.W., van Esch-Bussemakers, M.P. & Neerincx, M.A. (2006). "Designing personal attentive user interfaces in the mobile public safety domain." Computers in Human Behavior **22**(4): 749-770.

Van der Steur, G.A. (2016). “Analyse en aanpak van ziekteverzuim bij politie” [Analysis on and countermeasures to absence due to illness]. Ministerie van Veiligheid en Justitie, Den Haag, The Netherlands.

Van Es, A. & Stoker, E. (2015). “Onrust bij de politie: weinig vertrouwen in top” [Unrest at the police: low trust in management]. Volkskrant, 22 April 2015, http://www.volkskrant.nl/binnenland/onrust-bij-de-politie-weinig-vertrouwen-in-top~a3972486/

Van Loon, M. & Klumpenaar, S. (2015). “Politiebonden dreigen met nieuwe acties als fatsoenlijk bod uitblijft” [Police unions threaten with new actions in absence of an acceptable offer]. NRC, 16 September 2015, http://www.nrc.nl/nieuws/2015/09/16/actievoerende-agenten-beginnen-blokkade-ministeries.

Verwey, W.B. & Veltman, H.A. (1996). "Detecting short periods of elevated workload: A comparison of nine workload assessment techniques." Journal of Experimental Psychology: Applied **2**(3): 270-285.

Wickens, C.D. (1984). Processing resources in attention. In: Parasuraman, R. & Davies, D.R (Eds.), Varieties of attention. London, Academic Press**:** 63-102.

Wickens, C.D. (2008). "Multiple resources and mental workload." Human Factors **50**(3): 449-455.

Yang, C.-W., Yang, L.-C., Cheng, T.-C., Jou, Y.-T. & Chiou, S.-W. (2012). "Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room." Nuclear Engineering and Design **250**: 713-719.

Young, M.S., Brookhuis, K.A., Wickens, C.D. & Hancock, P.A. (2015). "State of science: mental workload in ergonomics." Ergonomics **58**(1): 1-17.

Zijlstra, F.R.H. (1993). Efficiency in work behavior - A design approach for modern tools. Delft, The Netherlands, Delft University of Technology. PhD thesis.

Chapter 2 has been published as:

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2016). Transitional Journey Maps: Reflections on creating workflow visualizations. In: Arezes, P.M. & Rodrigues de Carvalho, P.V. (Eds.) Ergonomics and Human Factors in Safety Management. CRC Press: 271-304.

Chapter 2

Transitional Journey Maps: Reflections on creating workflow visualizations

2.1 Introduction

Our daily lives are filled with interruptions and transitions from one task to another, resulting in a fragmented workflow. These can be students who knock on our doors when we are writing a paper, or traffic updates that require us to reschedule our route to work. Consider nurses who sequentially divide their attention between patients (e.g., Potter et al., 2004). Or consider a team of police officers, who just transported a suspect to the police station after a demanding pursuit. They are about to process the corresponding paperwork at their office when they receive an urgent call, after which they start driving to the reported incident location. The historical profiles of task transitions have been associated with recuperation in task performance (Matthews & Desmond, 2002) and mental workload (Morgan & Hancock, 2011). Furthermore, there is a substantial body of research that investigates the impact of interruptions on our work and well-being (e.g., Monk et al., 2008; Bailey & Iqbal, 2008). However, as

Baethge (2013) argues, these studies typically focus on isolated interruptions, thereby neglecting the accumulation of many interruptions throughout a day. As a result, she continues, an understanding of isolated interruptions cannot be generalized to a working day. In addition, Randall et al. (2000) argue that theoretical constructs based on findings in one domain may not be generalizable to another domain. These notions of limited ecological validity and generalizability have resulted in a move outside of the familiar laboratory environment, judged by the increasing amount of field studies in living labs (e.g., Keyson et al., 2013; Vastenburg et al., 2009; Niitamo et al., 2006). Changes in research methodology cause changes in the way we present and, consequentially, interpret our data. Data visualization facilitates exploration by transforming large amounts of textual or numeric data into graphical formats (Kondaveeti et al., 2012; Segelström, 2009; Card et al., 1999). Yet, to our knowledge, there are no guidelines regarding data visualization of workflows.

We were approached by two organizations with the request to study human information processing activities at work. The Dutch National Police was in the process of updating information technologies in their vehicles. They were interested in knowing how much information police officers can process in various work situations. This knowledge was to be translated into a set of requirements to aid in the selection of appropriate information technologies. Next, the European Space Operations Centre (ESOC, Darmstadt, Germany) wanted an improvement of the alarm sound design in their satellite control rooms. An evaluation of how operators deal with these signals in their workflow was used to inform the subsequent alarm design process. Although these contexts appear very different at first sight, the two case studies presented in this chapter show that both workflows are characterized by frequent task transitions and interruptions.

Our background in informational ergonomics was one of the reasons why we were approached. Informational ergonomics is about understanding how people use information, but also about understanding how to communicate information through design (e.g., visualizations). Thus, in both studies, workflow analyses were performed as input for subsequent research and design activities. Consequentially, the act of creating workflow visualizations became part of the design process.

Throughout our investigations, we encountered several theoretical and practical questions on how to interpret the data as function of categorization and visualization. The objective of this chapter is not to provide a final answer to all these questions. Rather, it is our hope that our way of dealing with these questions will foster critical reflection among those who wish to perform future studies on workflow-based information processing contexts.

2.1.1 Levels of abstraction

The problem of highly fragmented workflow lies in the fact that (1) one cannot finish an activity before a transition to another activity is required, and (2) it takes time to change one's mind-set back to the original activity (Monk et al., 2008). Zheng et al. (2010) define workflow fragmentation as the rate at which operators switch between tasks. Alternatively, González and Mark (2005) quantify workflow fragmentation as the average time continuously spent on an activity, before a transition takes place. In both cases, increased levels of workflow fragmentation are found at decreasing durations of activity segments. An important question from an information design perspective is when to best provide an operator with an information item. Since some activities typically last longer than others (e.g., reading vs. writing a paper), it makes sense to calculate workflow fragmentation separately for each activity category. This notion favors the time expenditure–based perspective on workflow fragmentation.

The next question, then, is at which abstraction level activities should be defined in order to measure their durations. We will explain the consequences associated with this question through an example of driving a car. Michon (1985) describes driving behavior on three levels: strategic, tactical, and control. The strategic level concerns general plans, such as route choice and scheduled destination time. The tactical level concerns planned activity patterns, such as overtaking and merging. Finally, the control level concerns automatic activity patterns, such as lane keeping and breaking. Figure 2.1 depicts transitions between activity categories over time. The strategic, tactical, and control levels are related to each other, in that driving a complete route at the strategic level encompasses a sequence of maneuvers at the tactical level (e.g., s1 consists of t1-t3-t4-t1), each of which in turn consists of sequential activity at the control level. Note that there is no one-to-one relationship between the levels; actions at the control level can be part of several maneuvers at the tactical level (e.g., steering actions can be found in overtaking, but also in merging).

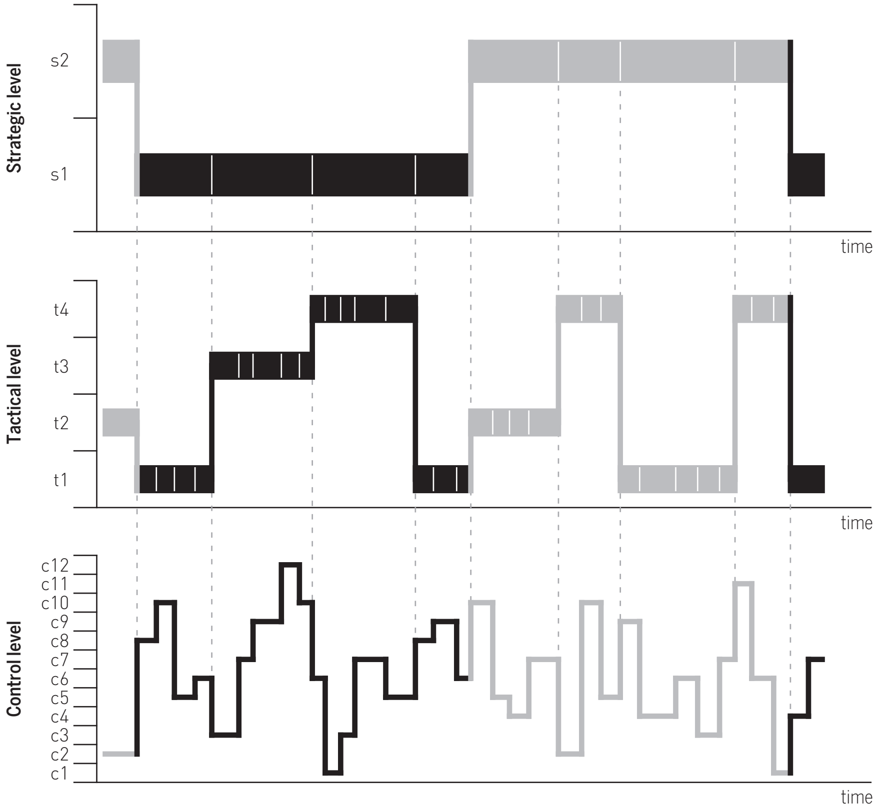

Figure 2.1. Three levels of driver behavior as introduced by Michon (1985): strategic (a), tactical (b), and control (c). In each panel, the horizontal axis represents time. Hypothetical data illustrate how segments within each level can be categorized on the vertical axis, and how clusters of segments on a lower level are the basis for transitions on higher levels.

The execution of a plan at the strategic level can take up to several hours, unless circumstances necessitate a change of plans (e.g., traffic updates). The duration of maneuvers on the tactical level take several seconds, whereas activities on the control level are described in terms of milliseconds. As a result, capturing all transitions at the control level requires a higher sample rate than those at the strategic level. Moreover, the example in Figure 2.1 shows how describing the same workflow at a higher level of abstraction results in longer activity segments, and less apparent workflow fragmentation. This notion raises a related question: Which level of abstraction results in a meaningful categorization of activities?

The relation between abstraction level, sample rate, and fragmentation not only is relevant for traffic research, but also should in fact be considered in any domain studied by the human factors, ergonomics, and human–computer–interaction communities. In these communities, two common frameworks to describe work at different levels of abstraction are activity theory (e.g., Nardi, 1995) and Rasmussen's (1983) abstraction hierarchy. Michon's (1985) strategic, tactical, and control driving behavior levels are comparable with, respectively, the working sphere (e.g., addressing purpose), action (e.g., goal) and operation (e.g., automatic condition) levels of activity theory as described by González and Mark (2005). Alternatively, they are related to the abstract function (e.g., addressing why), generalized function (e.g., what), and physical process (e.g., how) levels of the abstraction hierarchy. In the two case studies presented next, we have categorized workflow according to the overarching goal of the corresponding activities, which corresponds with the action or generalized function level. Data were collected through what eventually became Transitional Journey Maps, a new method to visualize workflow. By describing intermediate visualization stages, we will show that finding a meaningful level of abstraction can be the outcome of an interpretation process, rather than the starting point.

2.2 Case study 1: Dutch National Police vehicles [2]

The Dutch police force is currently looking for ways to improve the information system of their police vehicles, including pushing information (e.g., neighborhood updates and on-board training) to the vehicle. A central question is how much information officers can process in various work situations. Streefkerk et al. (2006) argued that a mobile police information system should be context-aware (i.e., involving time, location, environmental, and social factors) to prevent cognitive overload. This implies that the dynamics of police work should be taken into account for the development of such a system. For example, indications on average time spent on an activity and the corresponding mental workload may assist in determining the length and appropriateness of an information event (i.e., a moment during which information is presented). As it turns out, a detailed description of work dynamics is lacking in police literature. Therefore, the goal of this study is to better understand cognitive demands imposed on police officers by capturing the dynamics of operational policing.

2.2.1 Fragmentation in police work

Tromp et al. (2010) describe the work of Dutch police officers in terms of three activity categories: static, dynamic preventive, and dynamic reactive. In the static activity category, police officers are not assigned to a specific call, and they are working either at the office or in a parked vehicle. The dynamic preventive category concerns surveillance activities in a moving vehicle. Finally, police officers are said to operate in the dynamic reactive category when they are assigned to an urgent call while in their vehicle. Lundin and Nuldén (2007) identified five ways in which Swedish officers used their patrol car: on their way to an incident, on their way from an incident, at the site of an incident, for general surveillance when driving around or parked at a specific location, and parked at the station handling detained people or paperwork. A comparable categorization was found in

---

[2] The case study presented here is a revised version of Jansen et al. (2014).

a study on British police officers interacting with mobile technology (Sørensen & Pica, 2005). Here, the researchers distinguished five primary activity types: waiting in the car before an incident, driving to an incident, taking action at the incident, driving from the incident, and waiting in the car after an incident. Furthermore, they emphasized that this so-called generic cycle of operational policing can be interrupted and rearranged due to intermediate events (e.g., incoming calls with a higher priority). Borglund and Nuldén (2012) share this statement, identifying work rhythm as a problem area in the Swedish police force: "Much of police work is characterized by interruptions. Planned and ongoing activity can be discontinued at any time. Current routines and access to computer-based systems create a somewhat fragmented work situation for the officers" (p. 97). Similar accounts have been reported for the U.S. (Straus et al., 2010) and Dutch (Bouwman et al., 2008) police forces. Thus, the notion of fragmented work seems acknowledged in literature on operational policing.

Given the continuous switching between activities, it is important to not only focus on stationary mental workload during an activity, but also consider the effects of transitions between activities on mental workload. Yet, detailed investigations into police routines are typically represented through activity statistics using a full work shift as the time window (e.g., Anderson et al., 2005; Frank et al., 1997; Smith et al., 2001). These statistics do not provide information on whether an activity is executed without interruptions, or about patterns of fragmentation. Moreover, these investigations do not reflect police officers' subjective experiences related to these activities. While attempts to characterize police work fragmentation using scenarios (Borglund & Nuldén, 2012) or narratives (Sørensen & Pica, 2005) do include subjective experiences, they fail to quantify fragmentation. Therefore, the present study aims to unite a quantitative description of work dynamics with subjective experiences related to cognitive demands.

2.2.2 Method

A series of ride-alongs with Dutch police officers were arranged. Based on the method of contextual inquiry (Beyer and Holtzblatt, 1997), officers were interviewed and observed in their natural work environment, where they provided explanations as their work unfolded.

Participants

Ten officers (eight males, two females) volunteered to be accompanied in their patrol cars. Each officer had at least 2 years of experience with operational policing. Four ride-alongs were arranged, including three full 8-hour shifts and two shift changes in total. Hence, the vehicle was chosen as the central focus during ride-alongs, while personnel configurations changed from shift to shift. The ride-alongs included solo (two cases) and dual (four cases) patrol. With durations varying between 4.5 and 11 hours, in total 28 hours of data were collected. Colleagues of the officers often asked the researcher about his presence during stops at the police station. Their comments on work dynamics and organization are treated as part of the study results.

Apparatus

Data were collected with pen and paper, featuring timestamps, descriptions of the current activity, events in the officer's information environment that caused a transition to another activity (e.g., incoming calls and comments following an officer's observation), and utterances related to cognitive demands. All data were initially logged on a template with three rows of predefined activity categories. As requested by the client, these activity categories corresponded with the classification of Tromp et al. (2010) (i.e., static, dynamic preventive, and dynamic reactive).

Procedure

Before the ride-along began, the researcher explicitly stated that the study was not intended to judge the officers' performance. Agreements were made on safety and privacy. During the ride-alongs, the researcher tried to minimize hindrances by discretely observing what was going on. This nonparticipatory research approach was at times violated, for example, when an officer asked for details about a recent call. Existing studies recommend that the relationship with the officer should not be sacrificed for the sake of minimizing reactivity (Stol et al., 2004; Spano and Reisig, 2006). Interestingly, such a question can be regarded as a verbalization related to high cognitive demands. Officers were occasionally asked to explain what happened during transitions, but only if the work demands allowed for such concurrent reports. Otherwise, they were asked to give a retrospective report shortly after the event.

2.2.3 Results

A new method to visualize workflow will be introduced. The method is used to report findings on cognitive overload, and differences between solo and dual patrol.

Activity categorization

The left panel of Figure 2.2 displays the first page of the original field notes (in Dutch) of the first ride-along. The horizontal and vertical axes correspond with time and activity category, respectively. The text fields concern observations of and statements by a team of two police officers. The police officers were initially surveilling the neighborhood, until they were assigned to an incoming call. This was noted with "to incident" (Dutch: naar melding) in the dynamic reactive category. A few minutes later the call was cancelled by the dispatcher, as noted with "cancel" in the dynamic preventive category. Two arrows were drawn to connect the sequence of notes, thereby creating a sense of order and time. As a result, two transitions between the dynamic preventive and dynamic reactive activity categories were visualized. Next, an alarm sound (whiew) of the automatic license plate detector (Dutch: ANPR) was heard. One officer asked about the location of the detected car (Dutch: "Waar is-ie?"), to which the other officer replied that the car went in the opposite direction (Dutch: "Tegengestelde richting"). The officers' active search response was

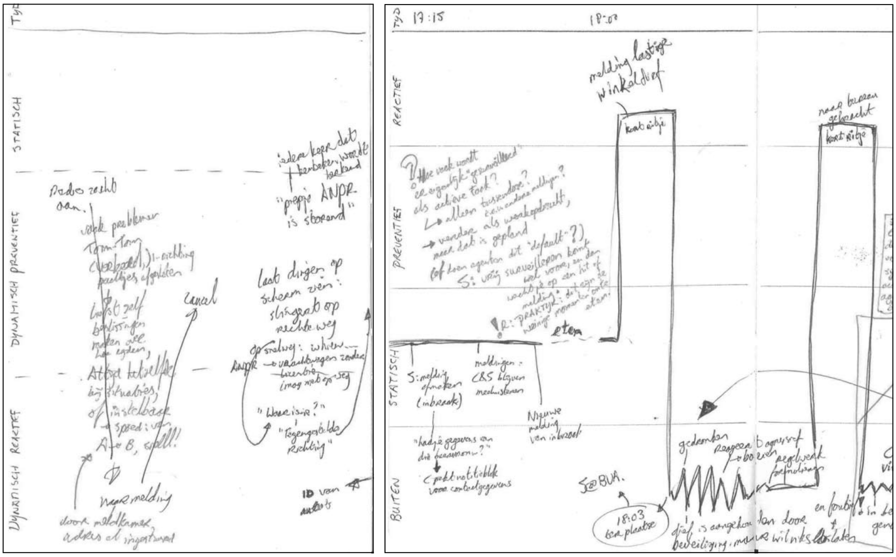

Figure 2.2. Original field notes of the first observation session (left panel) and the third observation session (right panel).

interpreted as a transition from the dynamic preventive to the dynamic reactive category. When it turned out that the detected car could not be intercepted, an arrow was drawn to indicate a transition back to the dynamic preventive category.

A section of the field notes of the third ride-along is shown in the right panel of Figure 2.2. Compared to the former field notes, there are differences in visualization style, the number of activity categories, and the arrangement of the activity categories. There is a continuous line that represents the activity category in which the police officers momentarily operated, and transitions between activity categories. This continuous line is augmented with text fields, whereas previously the text fields were augmented with arrows when there were transitions. Thus, the visualization style evolved as result of the perceived necessity to connect events in an orderly way. Furthermore, dynamics within activity categories were captured through different line styles. For example, a dashed line was drawn when the police officers were having a short break (Dutch: eten), and the aggressive behavior of a drunken shoplifter was represented through a zigzag line.

An additional activity category was introduced after finding a recurring activity that was not represented by the definitions of the existing categories. Although the static category covers situations in which police officers have left their vehicle, it does not include engaging at an incident. The capacity to interact with information technology is unlikely comparable between, for example, office work and handcuffing a drunken shoplifter. Such reactive behavior appears to be covered within dynamic reactive. However, that category does not include situations outside of the vehicle. Therefore, the outside (Dutch: buiten) category was introduced to represent police officers who have left their vehicle to engage at an incident (e.g., to catch the drunken shoplifter).

The order in which the activity categories were presented was changed twice. Whereas static, dynamic preventive, and dynamic reactive were originally visualized from top to bottom (see left panel of Figure 2.2), this order was reversed during the third ride-along (see right panel of Figure 2.2). This reordering was based on comments by police officers, who associated high driving speed levels during emergency situations with high adrenaline levels and low information processing capacities. Presenting the dynamic reactive activity category provided a better visual indication of the mental workload experienced by police officers.

The second rearrangement concerned the placement of the outside activity category. There was a logical reason why this activity category was originally presented at the bottom: the upper two categories were always related to a moving car, and the lower two categories were the only ones related to activities outside of the car. However, we observed that transitions to the outside category typically originated from the dynamic reactive and dynamic preventive categories. The visual appearance of the sudden drop from dynamic reactive to outside in the right panel of Figure 2.2 suggests that the workflow was disrupted, while capturing the shoplifter was actually a logical step after driving to the incident location.

In addition, many of these outside activities are likely associated with higher levels of mental workload than, for example, office work in the static category. Therefore, the outside activity category was eventually presented on top of the other categories.

Following a similar rationale as with the introduction of the outside category, the original framework of Tromp et al. (2010) was refined into six activity categories. Static was subdivided into "parked at the station" and "parked surveillance." Dynamic preventive was subdivided into "driving surveillance" and "driving to the station." Finally, dynamic reactive was subdivided into "driving to the incident" and "engaging at the incident" (formerly labeled "outside"). These six activity categories correspond with an adapted version of the framework by Lundin and Nuldén (2007), in that driving and nondriving surveillance activities were categorized separately.

Transitional Journey Maps

We refer to the graphical representation of interconnected objective data (e.g., observations) and subjective data (e.g., statements) as *Transitional Journey Map*. Four Transitional Journey Maps were constructed, one for each ride-along. An example can be found in the lower part of Figure 2.3. The vertical axis displays six activity categories, whereas time is found on the horizontal axis. The main actors are represented through three thick lines: the police vehicle (red violet), the driver (dark blue), and in the case of dual patrol, the co-driver (light blue). A journey through activity categories is created as the actors cross the underlying framework. Additional lines are used in case other actors come into play (e.g., the case of the copper thief, here represented in orange). Stationary vehicles are depicted with a dashed line. Similarly, dashed lines are used when officers are taking a break. Upon entering their car, officers' corresponding lines are joined with the vehicle's line. Segments of activities are demarcated by the time between adjacent transitions.

A transition is defined as a change from an activity category to another one. In Figure 2.3, transitions are labeled with hexagonal boxes, a character for

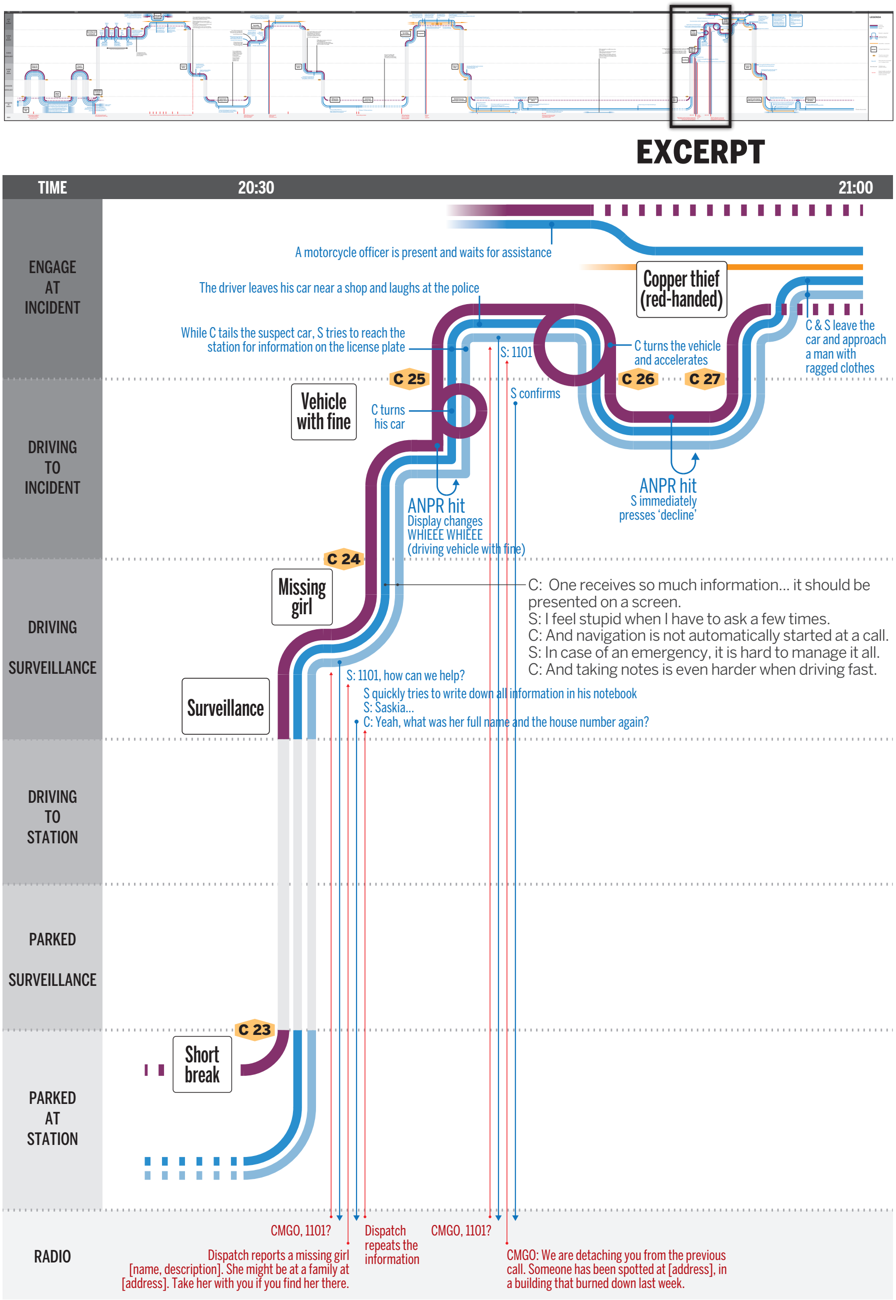

Figure 2.3. Excerpt (35 minutes) of a Transitional Journey Map (450 minutes). The horizontal and vertical axes display time and six activity categories, respectively. Additional details are described in the text.

the corresponding ride-along, and a number for the order of occurrence. For example, C23 refers to a segment of previous activity at the police station, and marks the transition from “parked at station” to “driving surveillance.” Descriptions for ongoing activities are depicted in white boxes for quick reference. Observations and utterances about an ongoing situation are depicted in a regular font style, whereas a an italic font style is used for retrospective accounts. Because of its dominant role in police work, instances of radio communication can be found in a separate row. The thin alternating arrows in Figure 2.3 show how messages are going back and forth between the officers and the dispatcher (e.g., the call of the missing girl).

Applying Transitional Journey Maps to operational policing

Visual inspection of a full Transitional Journey Map confirms the notion that police work is fragmented. The lower part of Figure 2.3 shows periods of many short activity segments, followed by relatively long stretches of paperwork at the police station. This is reflected in the boxplots of Figure 2.4, which show the durations of activity segments per activity category, including all ride-alongs. Outliers in the “engaging at incident” category were cases where victims or suspects were questioned, namely, theft (A7, C21) and domestic violence (B27, C13). All of these cases required more than half an hour of paperwork, with an outlier at 2 hours (C23). However, officers were often interrupted by incoming calls before finishing their office work, as reflected by the median duration of 17.9 minutes. Other outliers refer to picking up remote colleagues (A9), surveillance while bringing the researcher to the train station (Aend), and surveillance across a deserted national park (B8).

The categories “parked at station” and “driving surveillance” seem to take longer than “driving to station,” “driving to incident,” and “engaging at incident,” which seem to have similar segment durations (see Figure 2.3). Given the skewed distributions, nonparametric tests (SPSS v20) were used to compare between activity categories. As only one instance of parked surveillance occurred, this category was excluded from further analysis.

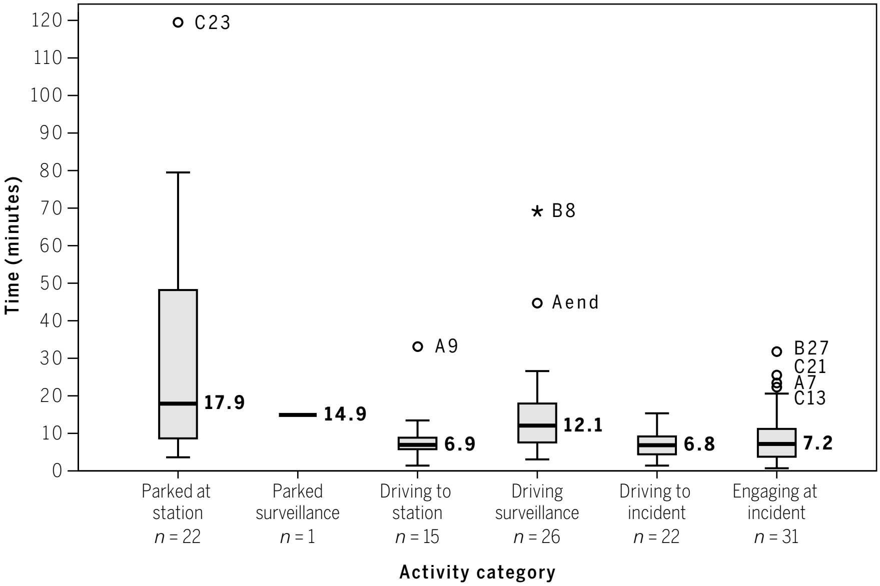

Figure 2.4. Boxplots of time spent in each activity category, summarized over all ridealongs. Median values are shown next to each box. Whiskers depict the lowest and highest data within a 1.5 interquartile range of the lower and upper quartiles, respectively. Outliers labeled with an asterisk or a circle concern solo and dual patrol, respectively.

Segment duration is significantly affected by activity category ($H(4) = 23.71$, $p < 0.001$). Seven Mann–Whitney tests were used to follow up this finding. Therefore, a Bonferroni correction was applied, and all effects are reported at a 0.007 level of significance. The duration of activities in the "parked at station" category was generally significantly longer than "driving to station" ($U = 70$, $r = -0.48$), "driving to incident" ($U = 90$, $r = -0.54$), and "engaging at incident" ($U = 148$, $r = -0.48$), but not longer than "driving surveillance" ($U = 208$, $r = -0.23$). Furthermore, activities performed in the "driving surveillance" category took significantly longer than "driving to incident" ($U = 139$, $r = -0.44$) and "engaging at incident" ($U = 235$, $r = -0.36$), but not longer than "driving to station" ($U = 108$, $r = -0.37$). It can be concluded that the most time for an informing event can be found when officers are working at the police station, or during surveillance while driving. Based on this dataset, an informing event should take less than 6.8 minutes if at least half of these events are to be fully processed in any activity category before a next transition takes place. However, these statistics do not address whether officers have spare capacity to successfully process the information.

Reports of cognitive overload

Comments by police officers regularly contained descriptions of situations witnessed during other ride-alongs, which were indicators for cognitive overload. For example, compare the following anecdote with Figure 2.3:

> *An incoming call instructs the officers to advance to a car that was broken into. L. takes a notebook from her pocket to record the address: "This way you don't have to ask again." A. responds: "On the group radio one often hears colleagues asking for a repetition of the suspect description. At the time they receive a call and they have to move as fast as possible, their mind set is already preoccupied."* (Field notes from ride-along 1)

Because of the activity descriptions and their characteristic visual pattern, the layout of a Transitional Journey Map facilitates remembering and retrieving events with related comments. Furthermore, the content of a comment dictates in which activity category it should be placed (e.g., a colleague at the station talking about an arrest belongs to "engaging at the incident"). Thus, an overview of information processing issues within an activity category can be obtained by scanning along the corresponding row in the Transitional Journey Maps. This approach resulted in the identification of an information processing paradox.

On the one hand, police officers not only monitor the radio for messages addressed to themselves, but also want to stay informed of the whereabouts and tasks of their colleagues. One reason is safety: *"If there is a call with violence, it's good to know if colleagues are nearby [...] then you know if and how long you should wait before stepping in."* Vice versa, officers may offer assistance. Second, there are functional implications: *"Those officers are busy over there, so I'll compensate by patrolling more centrally in this area."* Finally, it is part of a social system: when returning to the station after an emotionally demanding call, officers find support from colleagues that listened in. One officer commented that he was missing too much information, even though three channels were concurrently monitored (i.e., car radio and two earpieces).

On the other hand, police officers have trouble processing all information. As described above, incoming calls regularly contain too much information to remember. This is further inhibited by situational and state-related factors: *"If a situation is dangerous, you feel the adrenaline, stress, fatigue and tension, and this affects your ability to concentrate. In those situations, it is hard to hear something amidst other voices."* Messages are often hard to comprehend due to auditory masking by the police vehicle (e.g., when driving at high speed, often accompanied by a siren) and signal degradation in the communication system. In the meantime, the continuous monitoring and filtering of radio messages takes its toll. Up to 26 messages were counted in a time span of 5 minutes. Officers complained about high volumes, occasional feedback beeps, and fatigue: *"My left ear is deaf for other sounds because of the earpiece. After a busy shift, I still hear the voices at home."*

Comments on the necessity of monitoring radio communication were found in all activity categories, except for "parked surveillance." However, the representativeness of this exception is doubtful, since action in this category was observed only once. Comments on auditory masking were found in all activity categories that involved driving. Comments on overload were found in all activity categories, except for "driving to station." Overall, the observations and comments suggest that police officers want more information than they can handle with the current system.

Comparison of solo and dual patrol

All outliers in Figure 2.4 were cases of dual patrol, except for B8. This suggests a considerable difference in time spending between solo and dual patrol and, as a result, more time for information events during dual patrol.

Nonparametric tests were performed per category. Using a Bonferroni correction, the effects were compared with an alpha level of 0.008. None of the tests on time spending reached statistical significance. Nevertheless, police officers did mention differences between solo and dual patrol modes. The biggest impact is the opportunity to distribute tasks among officers in the case of dual patrol. Generally, the driver only concentrates on driving, whereas the co-driver is responsible for communication and surveillance tasks. Many officers commented that it is hard to operate the mobile data terminal while driving solo. Additionally, there are organizational differences between the patrol modes: *"If you're patrolling solo, you only get a call when the others cannot handle it. In cases of violence, we always operate with couples."* This suggests that differences may be found between the distributions of transitions.

Figure 2.5 depicts state diagrams for solo and dual patrol. An arrow line represents each cause for a transition between two activity categories. Thicker lines are used if the same cause was observed more than once. The total times spent observing solo and dual patrol were 12.6 and 15.1 hours, respectively. The relative time spent in each activity category is represented by the size of the corresponding circles. The two figures reveal that solo patrol involves relatively more driving surveillance activity than dual patrol (36% vs. 15%). Solo patrol involved more transitions from "driving surveillance" to "engaging at incident" (10 vs. 2), but less transitions from "driving to incident" to "engaging at incident" (5 vs. 12). Interestingly, in both patrol modes, 15 transitions were counted toward "engaging at incident." However, relatively more time on "engaging at incident" was spent in dual patrol (20% vs. 12%). This was caused by the longer times spent investigating incidents with violence (see outliers in Figure 2.2). Additionally, dual patrol involved more time spent on "driving to station" (12% vs. 3%), which may be due to the large amount of paperwork after serious incidents, and a higher likelihood of transporting victims or suspects afterward. In sum, the state diagrams on solo and dual patrol reflect the organizational differences uttered by the police officers.

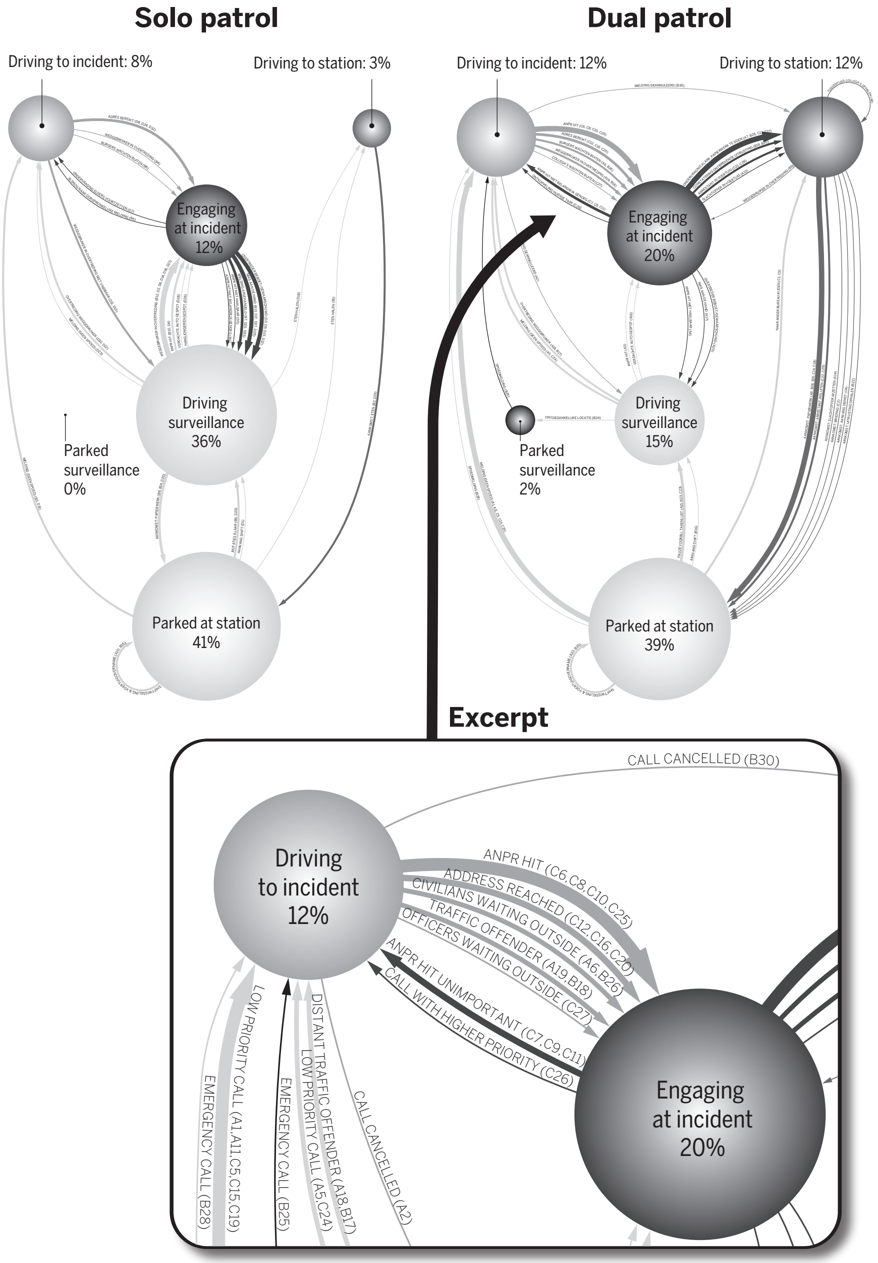

Figure 2.5. State diagrams of transitions during solo patrol (top left, 12.6 hours observed) and dual patrol (top right, 15.1 hours observed). The excerpt shows observed causes for transitions between activity categories. Codes in parentheses refer to the hexagonal boxes (i.e., transitions) in the Transitional Journey Maps.

2.2.4 Discussion

The information environment of a police officer is embedded in an unpredictable workflow, and includes numerous visual and auditory information channels. It has been argued that knowledge of work dynamics and related cognitive demands is beneficial for the development of an information and communication system. Regarding work dynamics, the current data suggest that in most cases, an information event that takes place while driving may not take longer than 6.8 minutes if it is to be fully processed before a next transition takes place. However, this finding does not address cognitive demands. The comments made by officers suggest that the demands of concurrent driving, surveillance, and monitoring force them into a permanent state of task-related effort (De Waard, 1996). Therefore, although no complete performance breakdowns were observed, the continuously experienced high workload was exhausting in the long run. "Driving to station" is the only activity category that may be used for additional information events, given the absence of comments on cognitive overload. However, the applicability would be limited during solo patrol. Overall, analyses of workflow fragmentation and officers' comments on cognitive overload suggest that the police vehicle's cockpit should be improved, before pushing additional information can be considered. Particularly, alleviation of the auditory channel is needed. The current results warrant further application of the Transitional Journey Map method in other contexts, such as information processing by other emergency services, control rooms, and tracking group dynamics for crowd management.

2.3 Case study 2: ESOC satellite control rooms [3]

The European Space Operations Centre (ESOC) requested us to evaluate and improve the auditory signals in their satellite control rooms. One of the issues was that warning beeps occur so often that they are ignored by the

---

[3] The case study presented here is a revised version of Jansen et al. (2015).

space controller (spacon). Also, irrelevant alarms disrupted the activities of colleagues working at a nearby workstation. There have been cases where the spacon would turn off those alarms, because their presence was causing too much stress. In both cases, the result is an increased risk of missing critical events.

These issues reflect characteristics associated with the safety hazard known as alarm fatigue. Alarm fatigue occurs when operators are continuously exposed to a large number of false alarms. When operators start to distrust these disruptive alarms (Cvach, 2012), they may disable, silence, or ignore them (Korniewicz et al., 2008). This has resulted in patient injury and death in the medical context (Sendelbach & Funk, 2013). The general strategy to counteract alarm fatigue is to reduce the number of unimportant alarms. For example, standards on alarm systems in process industries (ANSI/ISA 18.2, 2009; NEN-EN-IEC 62682, 2015) specify targets for average and peak alarm rates per operator console. These standards describe a rationalization stage, with the purpose of ensuring that alarms are only implemented if they are actionable (e.g., requiring a response).

However, there are potential side effects to alarm reduction. In communication between a satellite and ground control, a loss of signal is normally considered to be an abnormal condition, and should result in an alarm. Woods (1995) describes an example in which operators complained that the same alarm went off during each scheduled transponder switchover. When the system engineers removed the alarm, however, the operators complained that there was no indication that the event was occurring as expected. Rauterberg (1999) investigated the effect of auditory feedback on monitoring a simulated industrial plant with multiple machines. Each machine generated both auditory and visual feedback to inform the operator of its events. The auditory signals were redundant, as they referred to the same information as the visual signals. However, removal of the auditory signals resulted in decreased plant performance. Another study in the medical context shows that auditory signals may not always result in a response, but nurses use these signals as indicators for a patient’s status (Bitan et al., 2004). In sum, the removal of redundant signals does not necessarily improve performance, if these alarms provide meaningful information (Stanton et al., 2000).

Moreover, these examples show two facets of auditory signal interpretation: (1) operators anticipate system state changes within a certain context, and (2) alarm signals are sometimes used as feedback signals to confirm anticipated state changes. Therefore, alarm fatigue prevention may benefit from identifying which alarms are anticipated. The role of anticipation in alarm responses will be explored in the ESOC satellite control rooms.

2.3.1 Interpretation of auditory signals

In a typical control room, the system evaluates whether a set of process parameters are within specified limits. When a parameter crosses its limit, the system will generate a visual or auditory "out of limits" signal. Operators require contextual background knowledge to interpret the meaning of these signals (Seagull & Sanderson, 2001; Stanton, 1994). On the one hand, this knowledge determines whether the operator anticipated the event to which a signal corresponded. On the other hand, this knowledge allows the operator to determine whether the situation related to a signal is actionable. As a result, signals can be interpreted in four different ways, as shown in Table 2.1.

Table 2.1. Interpretations of an Auditory Signal, as a Function of the Actionability of a Situation, and Anticipation toward Events.

Situation	**Event**	
	Anticipated	**Unanticipated**
Actionable	Feedforward	Alarm
Nonactionable	Nuisance or feedback	Nuisance

The event state [actionable/nonactionable] separates actual alarms from signals that are generally experienced as nuisance. The situation state [anticipated/unanticipated] further refines these interpretations. A signal related to an anticipated event and an actionable situation will remind, rather than alert, that something is going to happen. Therefore, the related signal functions as feedforward.

Signals that relate to an anticipated event in a nonactionable situation can have two interpretations. In the first interpretation, they can be a nuisance. For example, an operator is informed that a core temperature crosses a threshold by hearing an auditory signal. However, this signal may also sound during maintenance on the system's ventilator, even though the temperature is temporarily allowed to cross the threshold. When the operator is aware of this situation, he or she may choose to ignore the signal. This illustrates how known malfunctions and planned system changes can turn an alarm signal into a nuisance signal. In the second interpretation, the example of Woods (1995) on loss of signal in communication between a satellite and ground control illustrates how an alarm signal is used as a feedback signal. This means that the value of anticipated nonactionable signals needs to be examined on a case-by-case basis.

2.3.2 System description satellite control rooms

ESOC accommodates several control rooms, each of which is related to one or more satellite missions. The distance to earth determines how long the contact with a satellite can be. This contact is referred to as a pass. All passes are scheduled. The dynamics of missions are different because of the period of contact and contact loss, as well as the distance between the satellite and earth. If the satellite is close to earth (mission type: earth observers), the pass duration, as well as the period between passes, is short. Contact with the satellite is almost instantaneous. Another option is a satellite with a fixed position in space (mission type: astronomy). This type of satellite is permanently in contact with the control rooms, but the antenna picking up the signal (antennas are located at three different places on earth) may change. If a satellite is at a long distance from earth (mission type: interplanetary), passes are long (e.g., 8 hours), but then the contact is also lost for a long time. Because of the long distance, it can take up to 20 minutes to send an instruction to the satellite, and an equal duration to receive a confirmation message from the satellite. Consequentially, there may be differences in anticipation toward events among the missions.

Each satellite is operated by a spacon, who monitors incoming data and events. When an error occurs, the spacon is informed by an auditory signal and an error message on the screen. For each mission, there is a specific protocol that the spacon has to follow. If necessary, an engineer is involved in resolving the problem.

2.3.3 Method

Three ESOC mission control rooms were visited (e.g., earth observer, astronomy, and interplanetary). The spacons on duty were interviewed and observed in their natural work environment, based on the method of contextual inquiry (Beyer & Holtzblatt, 1997). The primary focus was on auditory signals, as opposed to visual signals, given ESOC's request.

Participants

Four experienced male spacons were involved in this study. Each spacon was specialized in a single mission. In one mission, an experienced spacon was coaching an apprentice. In another mission, one spacon substituted for another spacon at the end of his shift.

Apparatus

Data were collected with pen and paper and two Roland R-05 portable field recorders. The notes consisted of timestamps, descriptions of the current activity, events in the spacon's information environment (e.g., auditory signals), utterances on work dynamics, and the presence of colleagues. The field recorders were primarily used to transcribe ongoing conversations between spacons, colleagues, and researchers, as well as voice loop communication.

Procedure

Spacons were informed about the presence of the researchers prior to the observations. After setting up the equipment, the researchers tried to minimize hindrance by discretely observing what was going on. Ongoing discussions between spacons and colleagues facilitated understanding of the situation. Spacons were occasionally asked to explain, for example, what they were working on, or what an auditory signal meant. These questions were only asked if the work situation allowed for such concurrent reports. Otherwise, spacons were asked to give a retrospective report shortly after the event. Engineers occasionally entered the satellite control rooms. Their comments on the ongoing mission status, as well as on auditory signals, are treated as part of the study results.

2.3.4 Results

In total, 6.6 hours of data were collected, during which 140 auditory signals were recorded (earth observer: 100 minutes, 38 signals; astronomy: 109 minutes, 34 signals; interplanetary: 185 minutes, 68 signals). Thirty-one auditory signals continued to ring until a spacon acknowledged them (e.g., ti-lu-li-ti-lu-li). Spacons labeled these sounds as alarms. Additionally, 109 signals were labeled as warnings (e.g., beep) or feedback signal (e.g., printer sounds). As these signals did not require an acknowledgment, it was not always possible to determine if they were actually heard by spacons. Therefore, only signals labeled as alarms are analyzed, unless stated otherwise.

The collected data of each mission are represented as a Transitional Journey Map. A state transition diagram is constructed to summarize transitions occurring in all missions. The two representations are used to perform a Bayesian inference on alarm anticipation, and to distinguish feedback signals from nuisance signals.

Activity categorization

The construction of a Transitional Journey Map requires a grouping of activities into a fixed number of activity categories. However, there was no documentation of activity categories at the start of this study. The first observation took place in the Cluster mission control room, where one out of four earth observer satellites could be controlled at a time. Our first

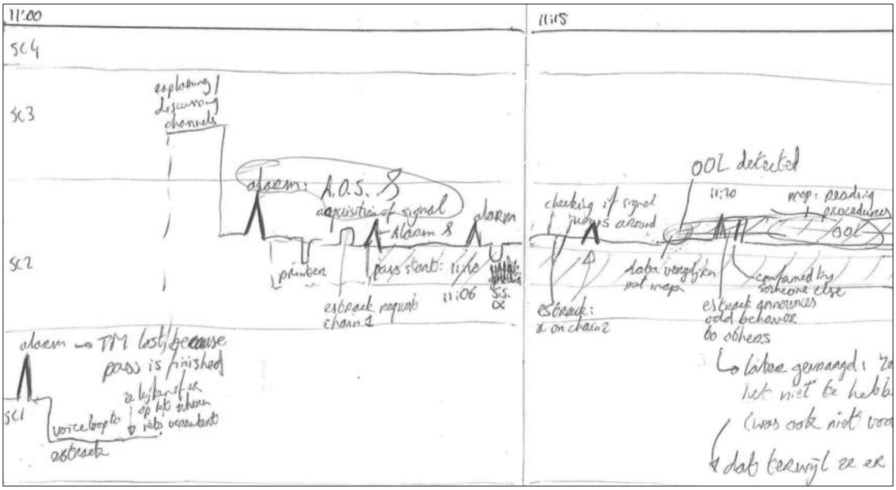

Figure 2.6. Original field notes of the Cluster mission control room. NOTE: Verbal data that were not transcribed during the observation session were later transcribed using audio recordings.

approach in categorizing the data was according to which satellite was momentarily controlled, analogous to how Potter et al. (2004) describe nurses' workflow as transitions from patient to patient over time. Figure 2.6 displays the field notes of the first observation. In this section, the spacons finished the pass of one satellite (e.g., sc1) and faced difficulties in establishing a connection with the next satellite (e.g., sc2). Auditory signals were represented as peaks in the continuous line. In some cases, the spacons expressed that there was no need to react to an alarm (e.g., "pass is finished"), whereas in other cases, alarms initiated problem-solving behavior (e.g., consulting protocols). Thus, there were a variety of responses to the same auditory signal. Unfortunately, the chosen categorization did not allow for visualizing this variety of responses in terms of transitions. Another problem was that no transitions were observed in the other missions, either because of the long pass duration (e.g., interplanetary) or because the mission did not feature passes (e.g., astronomy).

In our second approach, we compared the original system description with activities that were observed or described in all mission control rooms.

Four activity categories were derived. First, starting or ending a pass refers to activities such as requesting a new connection or breaking the existing connection. Second, monitoring can be characterized as a vigilance task: spacons monitor incoming data, ready to respond to potential problems. Third, commanding is about controlling the satellite, for example, by sending a list with maneuvers. Although this state involves monitoring activities as well (e.g., waiting for confirmations on the execution maneuvers), the purpose of this system state differs from that of the second system state. Fourth, solving problems involves activities related to unexpected messages (e.g., dealing with out of limits alarms).

Anticipation through Transitional Journey Maps

We used Transitional Journey Maps and a state transition diagram to represent the data required to fill out the cells of Table 2.1. Anticipation of events could not be measured directly, but was inferred from spacon statements. A transition between two activity categories was attributable to a signal if the former followed shortly after the latter. Also, it was not possible to observe directly whether the situation during which a signal occurred was actionable from the spacon's perspective. The presence or absence of a transition to an activity category of problem-solving behavior was used as a proxy for the actionability of a situation. This is in line with the purpose of an alarm, which is to inform a spacon that an abnormal condition occurred, which requires a response (ANSI/ISA 18.2, 2009).

Three Transitional Journey Maps were created. Figure 2.7 shows a 14-minute excerpt of the Cluster earth observer mission, which corresponds with a part of the data presented in Figure 2.6. The horizontal and vertical axes display time and the four activity categories, respectively. Transitions between activity categories are labeled with hexagonal boxes, which contain a reference to the mission and the order of occurrence (e.g., CL2). These lines are frequently interrupted by red circles, which represent auditory signals. A separate row for voice loop communication distinguishes between communication in the mission control room itself

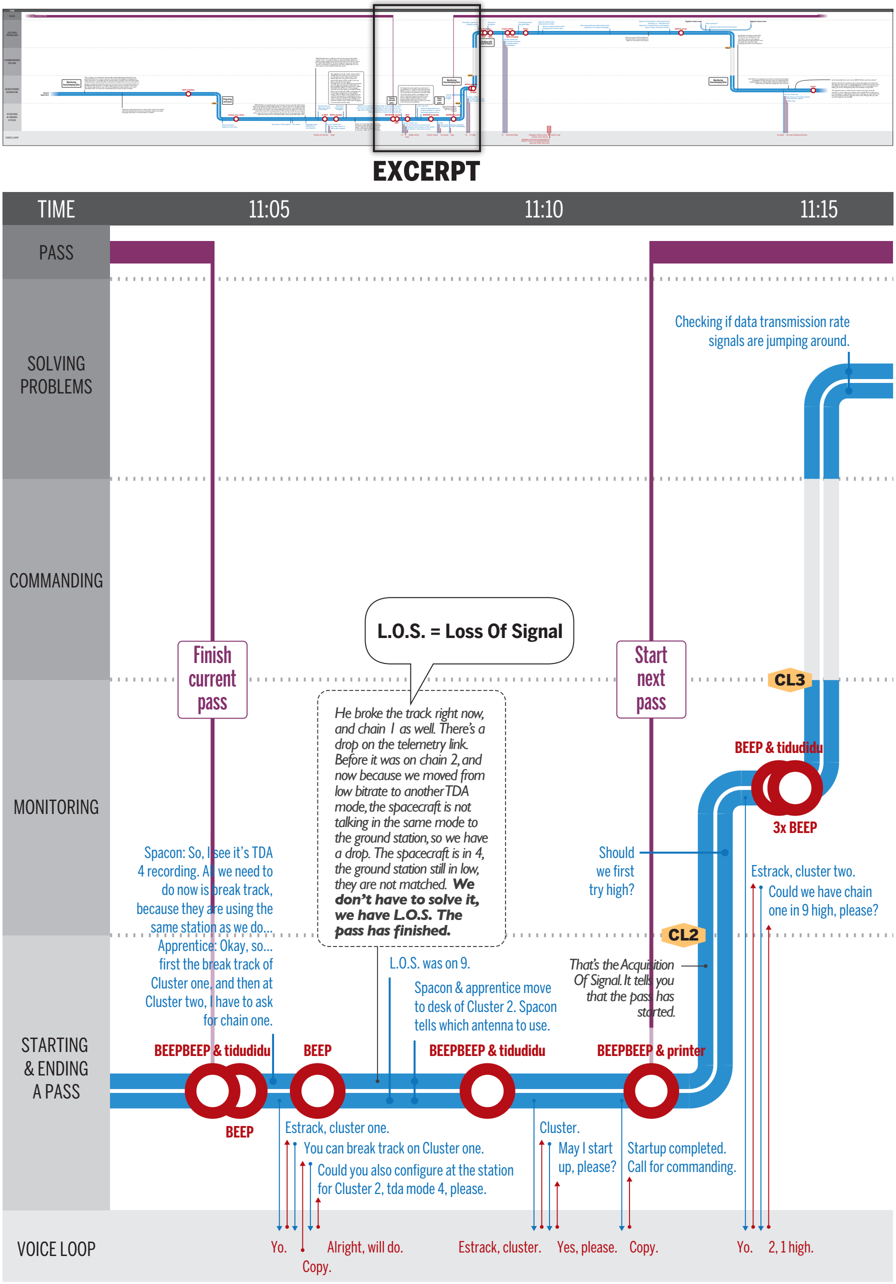

Figure 2.7. Excerpt (14 minutes) of a Transitional Journey Map (100 minutes). This excerpt describes the process of ending one pass and starting the next pass.

and communication with other rooms (e.g., engineers) or external parties (e.g., operators of the antennas, Estrack).

This excerpt illustrates the procedures between two passes. An example of anticipation can be found at the alarm at 11:05. This alarm is related to a loss of signal. The retrospective anecdote (regular font style) shows that the spacons knew this alarm would come. Additionally, anecdotes on and observations of the active situation (italic font style) suggest that this signal was used as a starting point for their activities toward the beginning of the next pass. In this case, the alarm sound did not initiate a transition toward problem-solving behavior, but functioned as a feedback signal. Another example of how spacons use auditory signals as feedback can be found at 11:12. The spacons knew that the next pass (e.g., "acquisition of signal") had started from the printer sound. This resulted in their transition (CL2) to the monitoring activity category. Finally, the alarm at 11:13 initiated a transition (CL3) to problem-solving behavior. The spacons determined that the signal-to-noise ratio of the data transmission was too low, and requested a switch to another bitrate. Their comment *"Should we first try high?"* indicates that they were prepared for this situation. These examples show how anticipation toward events can be derived from a Transitional Journey Map.

Alarm response behaviour through a state transition diagram

A state transition diagram was used to count and group signals that initiated problem-solving behavior. The Transitional Journey Maps were translated into a state transition diagram for an overview of transition causes and signal frequencies. In this diagram, each state corresponds with an activity category. Transitions were labeled according to the causes identified in the Transitional Journey Maps, and the frequency at which they occurred. A total of eight transitions between activity categories were found, which are represented as solid lines in the state diagram of Figure 2.8. The transition labeled CL2 in Figure 2.7 corresponds with the orange arrow from "starting and ending a pass" to "monitoring," whereas CL3 represents one of the critical events at the arrow from "monitoring" to

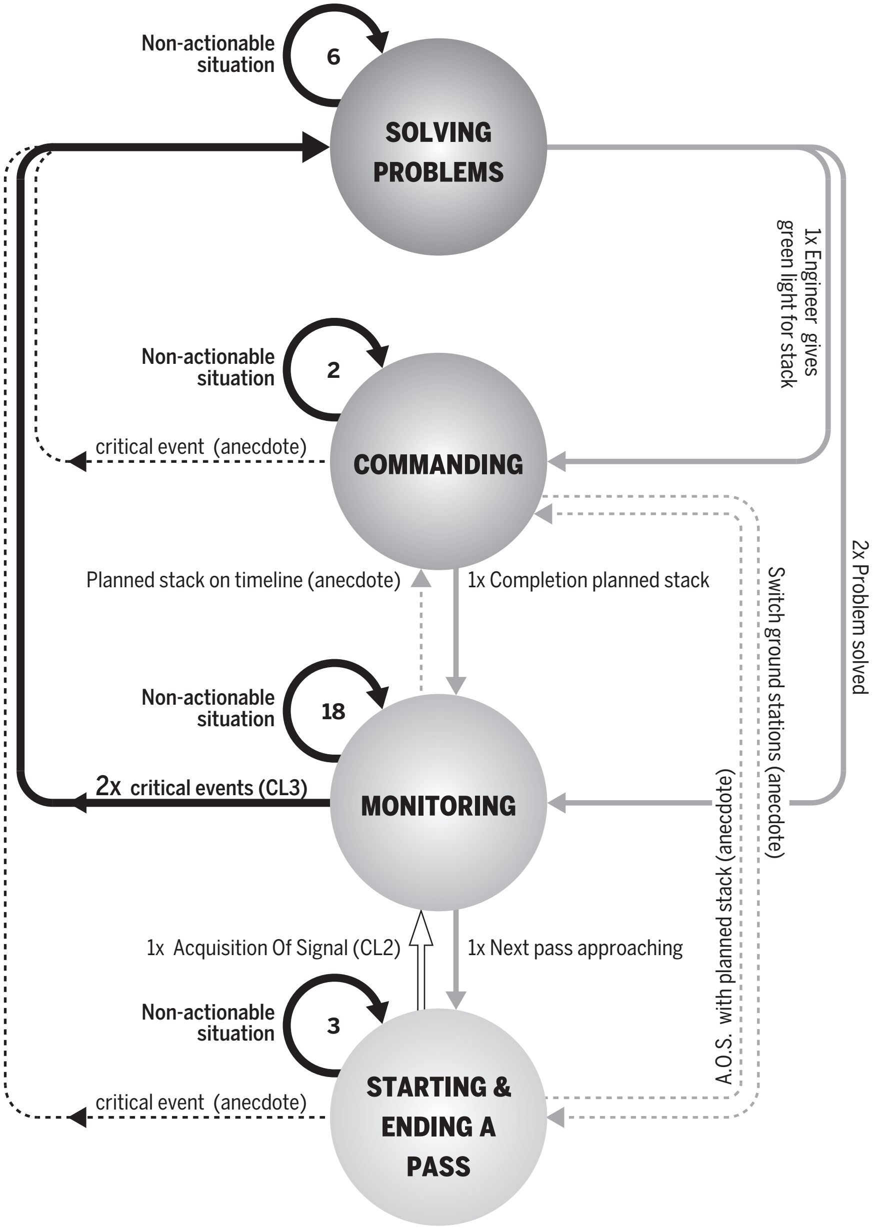

Figure 2.8. State transition diagram. Black and white lines represent state transitions initiated by alarm signals and printer sounds, respectively. Gray lines represent state transitions initiated by planned or finished activities. Solid lines are observed transitions. Dashed lines represent anecdotes.

"solving problems." Dashed lines represent transitions that were mentioned by spacons, but which were not observed. The state transition diagram shows that only two auditory signals labeled as an alarm resulted in a transition to problem-solving behavior. Statement analysis in the corresponding Transitional Journey Maps revealed that one of these signals came as a surprise, and therefore had an alarming function. The other signal was related to an anticipated event (e.g., transition CL3 described above).

The circular arrows in the state transition diagram show that 29 auditory signals labeled as an alarm did not result in a transition to another activity category at all, which implies they were nonactionable. For 19 of these nonactionable signals, statement analysis revealed that the corresponding system events were anticipated. The role of anticipation could not be derived for the other 10 signals. They occurred in situations where asking a spacon for an explanation would have disrupted his workflow, and where a retrospective explanation was hindered by other auditory signals that had rang in the meantime.

Quantifying the role of anticipation with Bayes' theorem

Bayes' theorem relates current probability to prior probability. Applied to the context of a control room, Bayes' theorem can be used to explore the causal relationship between event anticipation and situation actionability. Of particular interest in the present context is the probability that anticipation influenced spacons in deciding not to respond to a signal (e.g., the "nuisance or feedback" cell in Table 2.1). According to Westbury (2010), this study concerns the simplest form of Bayesian inference, with only two sets of mutually exclusive possibilities. Therefore, the canonical Bayesian expression can be rewritten as:

$$\begin{aligned} \text{P(Anticipated} \mid \text{Non-actionable)} &= \frac{\text{P(Non-actionable} \mid \text{Anticipated)} \bullet \text{P(Anticipated)}}{\text{P(Non-actionable)}} \qquad (2.1) \\ &= \frac{\text{P(Anticipated \& Non-actionable)}}{\text{P(Non-actionable)}} \end{aligned}$$

The two signals that initiated a transition to problem-solving behavior in the state transition diagram are represented in the "Actionable" row of Table 2.2. Signals corresponding with the circular arrows in the state transition diagram are found in the "Nonactionable" row of Table 2.2. As the role of anticipation could not be derived for 10 alarms, the distribution of nonactionable signals in Table 2.2 is represented as a range, where $0 \leq x \leq 10$. If all signals with unknown event anticipation were in fact unanticipated (e.g., $x = 10$), then Equation 2.1 yields a probability of 66%. A probability of 100% is found if all nonactionable signals were in fact anticipated (e.g., $x = 0$).

Table 2.2. Distribution of Alarm Sounds as a Function of Situation Actionability and Event Anticipation.

Situation	Event	
	Anticipated	**Unanticipated**
Actionable	1	1
Nonactionable	$19 + (10 - x)$	x

NOTE: The role of anticipation could not be derived for 10 alarms. See text for more explanation. Cells are divided by the total number of alarm sounds ($n = 31$) to obtain the probabilities in Equation 2.1.

Feedback signals versus nuisance signals

The Bayesian inference showed there is a high probability that spacons were guided by anticipation in their decision not to respond to signals. This warranted investigating whether the 19 nonactionable and clearly anticipated signals were interpreted as feedback or nuisance. Statement analysis in the Transitional Journey Maps resulted in the interpretations shown in Table 2.3. For 11 signals, the interpretation in terms of feedback value was unclear. Five signals were used to confirm the end of five unrelated processes (e.g., "loss of signal" in Figure 2.7). Therefore, these signals are marked as feedback signals. This was not the case for signals about a low temperature of the satellite. Spacons had disabled one of the heaters to save fuel, resulting in three nuisance signals.

Table 2.3. Interpretation of Anticipated Nonactionable Signals (19).

Activity category	**Feedback signal**	**Nuisance signal**	**Interpretation unclear**
Solving problems	0	0	0
Commanding	1	0	0
Monitoring	2	3	11
Procedures at pass	2	0	0

NOTE: All signals were designed as an alarm signal.

2.3.5 Discussion

This study shows that qualitative data analysis is essential for interpreting quantitative data on alarm responses. This was facilitated by the use of Transitional Journey Maps, by which alarm anticipation could be derived. Anticipation influenced spacons' choices to ignore auditory signals as alarms, which was confirmed through Bayesian inference. Expertise level may have induced anticipation, which allowed spacons to cope with high signal densities. This was found in all missions, which suggests that anticipation is not dependent on the type of mission (e.g., distance between satellite and earth), but rather a general strategy.

Only 2 of the 31 signals that were labeled as alarm actually corresponded with critical events. The rationalization stage in standards on alarm management (ANSI/ ISA-18.2, 2009; NEN-EN-IEC 62682, 2015) prescribes that all nonactionable signals are discarded. However, five nonactionable alarms related to anticipated system events were interpreted as feedback signals. We recommend designing these signals with a lower level of perceived urgency than the alarm signals. Concluding, the identification of valuable feedback signals should be part of the alarm management life cycle. This requires distinguishing between nuisance and feedback for anticipated nonactionable signals.

Most methods for alarm reduction target alarms with the highest frequency of occurrence (e.g., "bad actors") (Izadi et al., 2010) or groups of correlated alarms (Schleburg et al., 2013; Kondaveeti et al., 2012). These quantitative approaches do not take into account the context in which alarms were triggered. Consequentially, it is not possible to determine whether related events were anticipated, or if nonactionable signals contain valuable feedback information. However, a new monitoring paradigm with novelty detection (Martínez-Heras et al., 2012) may partially address this issue, for novelty excludes anticipation by definition. It is hoped that intelligent alarm systems will eventually reduce alarm fatigue by filtering nuisance signals, while continuing to provide valuable feedback.

2.4 General discussion

The Transitional Journey Map is a generic approach to describe workflow, connecting consecutive activities, occurring events, and verbal data as a journey along a timeline. Transitional journey maps can be used in varying contexts, ranging from rather unpredictable environments (e.g., incoming messages in operational policing) to reasonably predictable environments (e.g., anticipation in satellite control rooms). We discuss how the anatomy of a Transitional Journey Map enables three classes of analysis. Furthermore, we discuss why the act of creating data visualizations can be viewed as a design process, which justifies *post hoc* activity categorization.

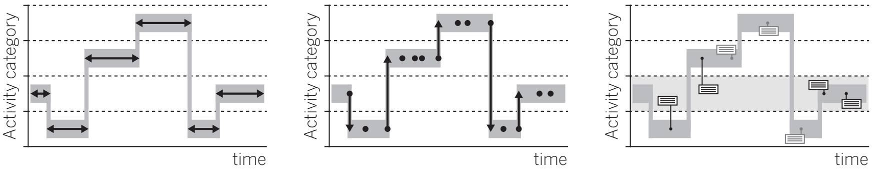

Figure 2.9. Analysis of a Transitional Journey Map based on segment lengths (a), transitions (b), and statements (c).

In Case Study 1, fragmentation in police work was quantified by calculating the average segment durations in each activity category (represented by horizontal arrows in Figure 2.9, left panel). This analysis demonstrates that the dynamics of operational policing are not properly represented by aggregating time expenditure per activity category over an entire shift (e.g., Anderson et al., 2005; Smith et al., 2001; Frank et al., 1997). Furthermore, differences between solo and duo patrol were examined through a comparison of their state transition diagrams. These diagrams were constructed through an inventory of all transitions between activity categories (see vertical arrows in Figure 2.9, center panel). Finally, opportunities for and limitations on human information processing were identified by clustering verbal data per activity category (see Figure 2.9, right panel). Apparently, police officers receive more information than they can process, yet they still desire to be informed about everything. This warrants a reconsideration of the current information system.

In Case Study 2, auditory signals in satellite control rooms were examined. Three Transitional Journey Maps were transformed into a state transition diagram, which showed that the majority of alarms did not result in a transition toward problem-solving behavior. The role of anticipation in decision making was identified, based on verbal data related to events that did result in a transition, and events that did not (see dots in Figure 2.9, center panel). Bayesian inference confirmed that anticipation played a role in spacons' decisions to ignore auditory signals.

2.4.1 Transitional Journey Maps as a result

The analyses described in Figure 2.9 are possible because of the anatomy of a Transitional Journey Map. Its main elements are the outline (e.g., x-axis and y-axis), the unit of analysis (e.g., actors, journey, and overarching system), and the interactions (e.g., observations and verbal data). In some studies, the time axis is categorized according to a predefined sequence of activities, such as customer journey maps (e.g., Trischler & Zehrer, 2012; Zomerdijk & Voss, 2010) and anesthesia procedures (e.g., Kennedy et al., 1976). However, the unpredictable aspects of police work and satellite control room operations refute predefined sequences. Therefore, no categorization is used on the x-axis of a Transitional Journey Map.

A transition between two activity categories should always signify disruption. To ensure meaningful transitions with regard to workflow fragmentation, activities on the y-axis of a Transitional Journey Map are categorized based on a common goal. When activities are categorized on a lower level of abstraction, there is a chance that a sequence of transitions is unintentionally associated with a high level of workflow fragmentation. For example, Zheng et al. (2010) assigned the acts of walking and talking (e.g., what) to separate categories, without reference to a goal (e.g., why). If a transition between these categories is part of the same momentary workflow goal (e.g., walking to a colleague to discuss something), it is illogical to state that the workflow was disrupted.

The combination of verbal data in a visualization helps in understanding human interactions with information systems. Previous studies have identified high levels of workflow fragmentation by showing transitions between activity categories. However, these studies stated that verbal data were required to determine the cause for these transitions (e.g., Cornell et al., 2010; Zheng et al., 2010; Potter et al., 2004). By including verbal data, Transitional Journey Maps enable the examination of such causality. In Case Study 1, for example, incoming messages from the dispatcher made it possible to determine the cause for transitions toward the “driving to an incident” activity category. In Case Study 2, absent reactions to auditory signals were explained by verbal statements in close temporal proximity of the occurring signals. In addition, Transitional Journey Maps incorporate the context in which statements are made. For example, the momentary communication state of the satellite control system was represented as a pass layer. The interpretation of verbal statements was facilitated by knowing that a pass had just ended. Finally, a distinction between retrospective and momentary verbal data facilitates triangulation between reported and observed events. Police officers in Case Study 1 gave retrospective accounts on differences between solo and dual patrol (e.g., police officers operate in couples in case of violence). This difference was confirmed by an event in another observation session (e.g., the motorcyclist waiting for assistance in Figure 2.3).

2.4.2 Transitional Journey Maps as a process

There were several iterations of data visualization during the data collection phase. This iterative process facilitated articulating insights during the analysis phase (cf. Segelström, 2009). The field notes presented in Case Studies 1 and 2 (see Figures 2.2 and 2.6) display early versions of the Transitional Journey Map. Initially, data were collected as a set of transcriptions in the appropriate activity category. Later, arrows between the notes provided a sense of order and time, and revealed transitions between activity categories. There were several changes in the number of activity categories, and the order in which they were presented. The latter was guided by the observation that some activity categories were associated with higher levels of mental workload. This transformed a nominal categorization into a semiordinal categorization.

This iterative process shows a strong resemblance with the sketching phase of a product design process. Ferguson (1992) distinguishes three sketch types. So-called thinking sketches support the thinking process of the individual designer, whereas talking sketches support group discussion within the design team, and prescriptive sketches are used to communicate detailed information outside the design time. Goldschmidt (1991) describes thinking sketches in terms of an interpretative cycle. A designer will not "see" the entire image in his or her mind before putting a preliminary version of this image on paper. It is through the act of sketching that the image is brought into existence, both on paper (e.g., the sketch) and in the designer's mind (e.g., knowledge). The acquired knowledge may serve as an inspiration for creating another sketch. A similar interpretative cycle appears to have occurred in creating Transitional Journey Maps. Furthermore, designers draw the same product from different angles, to explore which perspective best communicates relevant product properties. An analogy is found in the rearrangement of activity categories in Case Study 1. Finally, the original field notes comply with Buxton's (2007) characterization of sketches. In his view, sketches are suggestive and tentative, rather than descriptive and specific. Their aim is to explore, rather than to define. In sum, data visualization can be viewed as design activity in explorative field studies.

2.4.3 Implications

In the introduction of this chapter, a methodological question was raised about which level of abstraction results in a meaningful categorization of activities. We have argued that categorizing activities according to a common goal ensures meaningful transitions with regard to workflow fragmentation. However, in Case Study 1, the number of activity categories was extended within the same level of abstraction. Cornell et al. (2010) argue that a large set of activity categories enables one to detect subtle effects, but the downside of such a large set is the risk of losing one's overview of the larger workflow picture. In their study, previous research was used to establish a set of activity categories. An important finding of the present studies is the notion that data visualization is part of an ongoing sense-making process, which does not necessitate *a priori* categorization. Thus, in explorative field studies, the exact number of activity categories can also be determined during and after collecting the data.

Further explorations in workflow visualization should ideally collect data at multiple levels of abstraction. We envision a tool that reveals different workflow patterns by zooming in or out to a given level of abstraction, and by dynamically redefining the number of activity categories within that level. Because lower levels of abstraction require a higher sample resolution, an automatic logging system may help in gathering more detailed work patterns. We have shown that qualitative data are essential to understand quantified behavior. The challenge, then, lies in automatically capturing verbal data. While time-consuming, the presence of a researcher offers the opportunity to ask explanations about what is happening, and to clarify previous statements.

Transitional journey map visualizations combine quantitative data with qualitative data. This combination affords a better understanding of human interaction with information systems in a dynamic context. We hope this understanding will inform studies on workflow fragmentation and inspire the design of future information systems.

Acknowledgments

Case Study 1 was partially funded by Staf Korpsleiding/Directie Facility Management, Nationale Politie. Case Study 2 was partially funded by the European Space Agency (project code AO 1–7223/12/F/MOS). The authors wish to thank Sacha Silvester for his contribution to Case Study 1, and all police officers and spacons who volunteered for their cooperation.

References

Anderson, G.S., Courtney, A., Plecas, D., & Chamberlin, C. (2005). Multitasking behaviors of general duty police officers. *Police Practice and Research 6*(1): 39-48.

ANSI/ISA (2009). Management of alarm systems for the process industries. *18.2-2009*.

Baethge, A. (2013). *A daily perspective on work interruptions*. Sozialwissenschaften, Medien und Sport, Mainz, Johannes Gutenberg-Universität. PhD: 171.

Bailey, B.P., & Iqbal, S.T. (2008). Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management. *ACM Transactions on Computer-Human Interaction 14*(4): 1-28.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: defining customer-centered systems*, Morgan Kaufmann.

Bitan, Y., Meyer, J., Shinar, D., & Zmora, E. (2004). Nurses' reactions to alarms in a neonatal intensive care unit. *Cognition, Technology & Work 6*(4): 239-246.

Borglund, E., & Nuldén, U. (2012). Personas in Uniform: Police Officers as Users of Information Technology. *Transactions on Human-Computer Interaction 4*(2): 92-106.

Bouwman, H., Haaker, T., & De Vos, H. (2008). Mobile applications for police officers. *BLED 2008 Proceedings*, 78-90, Bled, Slovenia.

Buxton, W. (2007). *Sketching User Experiences: Getting the Design Right and the Right Design*. San Fransisco, Morgan Kaufmann.

Card, S.K., Mackinlay, J.D., & Shneiderman, B., Eds. (1999). *Readings in information visualization: using vision to think*, Morgan Kaufmann Publishers Inc.

Cornell, P., Herrin-Griffith, D., Keim, C., Petschonek, S., Sanders, A.M., D'Mello, S., Golden, T.W., & Shepherd, G. (2010). Transforming nursing workflow, Part 1: The chaotic nature of nurse activities. *Journal of Nursing Administration 40*(9): 366-373.

Cvach, M. (2012). Monitor Alarm Fatigue: An Integrative Review. *Biomedical Instrumentation & Technology 46*(4): 268-277.

De Waard, D. (1996). *The measurement of drivers' mental workload*. Traffic Research Center VSC. Groningen, University of Groningen. PhD.

Ferguson, E.S. (1992). *Engineering and the mind's eye*. MIT Press, Cambridge, MA.

Frank, J., Brandl, S.G., & Watkins, R.C. (1997). The content of community policing: a comparison of the daily activities of community and 'beat' officers. *Policing: An International Journal of Police Strategies & Management 20*(4): 716-728.

Goldschmidt, G. (1991). The Dialectics of Sketching. *Creativity Research Journal 4*(2): 123-143.

González, V.M., & Mark, G. (2005). Managing Currents of Work: Multitasking Among Multiple Collaborations. In: Gellersen, H., Schmidt, K., Beaudouin-Lafon, M., & Mackay, W. (Eds.) *ECSCW 2005*. Springer Netherlands: 143-162.

Izadi, I., Shah, S.L., & Tongwen, C. (2010). Effective resource utilization for Alarm Management. *49th IEEE Conference on Decision and Control (CDC)*, December 15-17, 2010.

Jansen, R.J., R. Van Egmond, & H. De Ridder. 2015. No alarms and no surprises: How qualitative data informs Bayesian inference of anticipated alarm sounds. *Procedia Manufacturing 3*:1750–1757.

Jansen, R.J., Van Egmond, R., De Ridder, H., & Silvester, S. (2014). Transitional Journey Maps: Capturing the dynamics of operational policing. In: De Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W., & Toffetti, A. (Eds.) *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*, 15-27, Torino, Italy.

Kennedy, P.J., Feingold, A., Wiener, E.L., & Hosek, R.S. (1976). Analysis of tasks and human factors in anesthesia for coronary artery bypass. *Anesthesia and Analgesia 55*(3): 374-377.

Keyson, D., Al Mahmud, A., & Romero, N. (2013). Living Lab and Research on Sustainability: Practical Approaches on Sustainable Interaction Design. In: Augusto, J., Wichert, R., Collier, R., Keyson, D., Salah, A., & Tan, A.-H. (Eds.) *Ambient Intelligence*. Springer International Publishing. 8309: 229-234.

Kondaveeti, S.R., Izadi, I., Shah, S.L., Black, T., & Chen, T. (2012). Graphical tools for routine assessment of industrial alarm systems. *Computers & Chemical Engineering 46*: 39-47.

Korniewicz, D.M., Clark, T., & David, Y. (2008). A National Online Survey on the Effectiveness of Clinical Alarms. *American Journal of Critical Care 17*(1): 36-41.

Lundin, J., & Nuldén, U. (2007). Talking about tools – investigating learning at work in police practice. *Journal of Workplace Learning 19*(4): 222-239.

Martínez-Heras, J.-A., Donati, A., Kirsch, M.G.F., & Schmidt, F. (2012). New Telemetry Monitoring Paradigm with Novelty Detection. *SpaceOps 2012 Conference*, 1-9, American Institute of Aeronautics and Astronautics, Stockholm, Sweden.

Michon, J. (1985). A Critical View of Driver Behavior Models: What Do We Know, What Should We Do? In: Evans, L., & Schwing, R. (Eds.) *Human Behavior and Traffic Safety*, 485-524. Plenum Press, New York.

Monk, C.A., Trafton, J.G., & Boehm-Davis, D.A. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Applied Experimental Psychology 14*(4): 299-313.

Morgan, J.F., & Hancock, P.A. (2011). The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human Factors 53*(1): 75-86.

Nardi, B. (Ed.) (1995). *Context and consciousness: activity theory and human-computer interaction*, Massachusetts Institute of Technology, Cambridge, MA.

NEN-EN-IEC (2015). Management of alarm systems for the process industries. *62682:2015*.

Niitamo, V.-P., Kulkki, S., Eriksson, M., & Hribernik, K.A. (2006). State-of-the-Art and Good Practice in the Field of Living Labs. *Proceedings of the 12th International Conference on Concurrent Enterprising: Innovative Products and Services through Collaborative Networks*. Milan, Italy: 349-357.

Potter, P., Boxerman, S., Wolf, L., Marshall, J., Grayson, D., Sledge, J., & Evanoff, B. (2004). Mapping the nursing process: a new approach for understanding the work of nursing. *Journal of Nursing Administration 34*(2): 101-109.

Randall, M.J., Roth, E.M., Vicente, K.J., & Burns, C.M. (2000). There Is More to Monitoring a Nuclear Power Plant than Meets the Eye. *Human Factors 42*(1): 36-55.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics 13*(3):257-266.

Rauterberg, M. (1999). Different Effects of Auditory Feedback in Man-Machine Interfaces. In: Stanton, N.A., & Edworthy, J. (Eds.) *Human Factors in Auditory Warnings*, 225-242. Ashgate, Sydney.

Schleburg, M., Christiansen, L., Thornhill, N.F., & Fay, A. (2013). A combined analysis of plant connectivity and alarm logs to reduce the number of alerts in an automation system. *Journal of Process Control 23*(6): 839-851.

Segelström, F. (2009). Communicating through Visualizations: Service Designers on Visualizing User Research. *First Nordic Conference on Service Design and Service Innovation*, 175-185, Oslo, Norway.

Sendelbach, S., & Funk, M. (2013). Alarm Fatigue: A Patient Safety Concern. *AACN Advanced Critical Care 24*(4): 378-386.

Smith, B.W., Novak, K.J., & Frank, J. (2001). Community Policing and the Work Routines of Street-Level Officers. *Criminal Justice Review 26*(1): 17-37.

Sørensen, C., & Pica, D. (2005). Tales from the police: Rhythms of interaction with mobile technologies. *Information and Organization 15*(2): 125-149.

Spano, R., & Reisig, M.D. (2006). "Drop the clipboard and help me!": The determinants of observer behavior in police encounters with suspects. *Journal of Criminal Justice 34*(6): 619-629.

Stanton, N. (Ed.) (1994). Alarm initiated activities. *Human factors in alarm design*, 93-117, Taylor & Francis, London.

Stanton, N.A., Harrison, D.J., Taylor-Burge, K.L., & Porter, L.J. (2000). Sorting the Wheat from the Chaff: A Study of the Detection of Alarms. *Cognition, Technology & Work 2*(3): 134-141.

Stol, W.P., van Wijk, P., Vogel, G., Foederer, B., & van Heel, L. (2004). *Politiestraatwerk in Nederland: Noodhulp en Gebiedswerk* [Police streetwork in The Netherlands]. Apeldoorn / Zeist, Nederlandse Politie Academie / Kerckebosch.

Straus, S., Bikson, T., Balkovich, E., & Pane, J. (2010). Mobile Technology and Action Teams: Assessing BlackBerry Use in Law Enforcement Units. *Computer Supported Cooperative Work (CSCW) 19*(1): 45-71.

Streefkerk, J.W., van Esch-Bussemakers, M.P., & Neerincx, M.A. (2006). Designing personal attentive user interfaces in the mobile public safety domain. *Computers in Human Behavior 22*(4): 749-770.

Trischler, J., & Zehrer, A. (2012). Service design: Suggesting a qualitative multistep approach for analyzing and examining theme park experiences. *Journal of Vacation Marketing 18*(1): 57-71.

Tromp, N., van Dijk, M., & Hekkert, P. (2010). *Toekomstig Mobiel Werken: een visie vanuit het sociale systeem* [Future mobile workplaces: a vision based on the social system]. Internal report Delft University of Technology, Delft, The Netherlands.

Vastenburg, M.H., Keyson, D., & De Ridder, H. (2009). Considerate home notification systems: A user study of acceptability of notifications in a living-room laboratory. *International Journal on Human-Computer Studies 67*: 814-826.

Westbury, C.F. (2010). Bayes' rule for clinicians: an introduction. *Frontiers in Psychology 1*: 1-7.

Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics 38*(11): 2371-2393.

Zheng, K., Haftel, H.M., Hirschl, R.B., O'Reilly, M., & Hanauer, D.A. (2010). Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective. *Journal of the American Medical Informatics Association 17*(4): 454-461.

Zomerdijk, L.G., & Voss, C.A. (2010). Service Design for Experience-Centric Services. *Journal of Service Research 13*(1): 67-82.

Chapter 3 has been published as:

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2016). Task Prioritization in Dual-Tasking: Instructions versus Preferences. PLoS ONE 11(7): e0158511.

Chapter 3

Task prioritization in dual-tasking: instructions versus preferences

Abstract

The role of task prioritization in performance tradeoffs during multi-tasking has received widespread attention. However, little is known on whether people have preferences regarding tasks, and if so, whether these preferences conflict with priority instructions. Three experiments were conducted with a high-speed driving game and an auditory memory task. In Experiment 1, participants did not receive priority instructions. Participants performed different sequences of single-task and dual-task conditions. Task performance was evaluated according to participants' retrospective accounts on preferences. These preferences were reformulated as priority instructions in Experiments 2 and 3. The results showed that people differ in their preferences regarding task prioritization in an experimental setting, which can be overruled by priority instructions, but only after increased dual-task exposure. Additional measures of mental effort showed that performance tradeoffs had an impact on mental effort.

The interpretation of these findings was used to explore an extension of Threaded Cognition Theory with Hockey's Compensatory Control Model.

3.1 Introduction

> *Police officer A. reflects on an incoming radio message: 'During an emergency call one receives a lot of information in a short timeframe. Such a call may include the shop name, crime type, potential dangers, suspect descriptions, which colleagues are on the case, and the plan. Meanwhile, you have to pay attention to the road, so sometimes you do not hear everything.' His colleague comments on the imposed organizational demands: 'In case of solo patrol you have to be much sharper [...] but you will commit so many traffic violations.'* (field notes in Jansen et al., 2014)

This example illustrates a common situation in our daily lives, namely, that we are asked to perform several tasks at the same time. This multi-tasking, however, often requires too much attention resulting in a conflict referred to as task interference (Bootes & Chapparo, 2010; Caird et al., 2008; Dressel & Atchley, 2008; Hembrooke & Gay, 2003). The obvious way to cope with task interference is to prioritize one task over the others (Gopher et al., 1989). But, as the police officers in the example show, this allocation of attention to one task goes at the expense of other tasks (Gopher & Navon, 1980; Norman & Bobrow, 1975). A possible solution was recently suggested by Salvucci & Taatgen in the form of continuous rapid switching between concurrent tasks (Salvucci & Taatgen, 2008). Over time this will yield the impression that these tasks are performed simultaneously and hence reported as multi-tasking.

In this paper, we claim that the concept of rapid switching between concurrent tasks needs an extension in order to accommodate another aspect of the example with the police officers, namely, that they seem to have different preferences in task prioritization. The first police officer missed incoming radio messages because he preferred to prioritize the driving task while the other police officer committed traffic violations as a result of paying more attention to the radio messages. This suggests that people have internal preferences regarding task prioritization. The role of preference on task prioritization has received limited attention. It is typically assumed that task prioritization can be obtained by means of an external priority instruction on the relative importance of each task (Dressel & Atchley, 2008; O'Donnell & Eggemeier, 1986). However, people are not always able or willing to follow priority instructions (Levy & Pashler, 2008; Miller & Durst, 2014; Nijboer et al., 2013; Siu & Woollacott, 2007). Cnossen et al. (2004) argue that judgments on performance decrements should be based on how people decide to prioritize between tasks, instead of what they are instructed to do. In order to understand what really happened in these studies, we first need to know whether preferences do exist and whether they may have an impact on the effectivity of task priority instructions. The question thus becomes: is there a possibility that when people are instructed to prioritize one task over another, but in fact prefer to perform the other task, they act according to their preference?

The aim of the present study is to provide an answer to these questions by performing a series of experiments in which participants had to perform two concurrent tasks. The first step is to verify whether people have preferences (Experiment 1). The second step is to focus on possible interactions between preferences and instructions (Experiments 2 and 3). The findings of this quasi-experimental study called for a theoretical exploration. Therefore, as a third step, we extended Salvucci & Taatgen's (2008) Threaded Cognition Theory with Hockey's (1997; 2011) Compensatory Control Theory as a representation of cognitive-energetic models on task performance. But first, we introduce the mechanisms of task interference and task prioritization as predicted by Threaded Cognition Theory.

3.1.1 Mechanism of task interference

Two tasks are said to interfere when simultaneous task execution results in decreased performance on one or both tasks (e.g., tradeoffs between missed

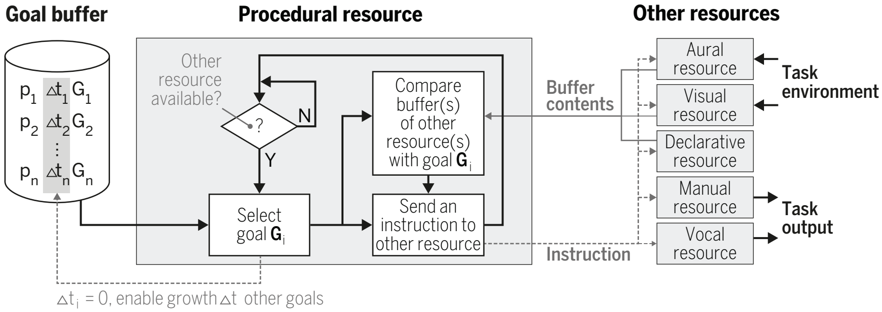

Figure 3.1. Control flowchart interpretation of Threaded Cognition Theory (Salvucci & Taatgen, 2008), with goal-related instructions fired by the procedural resource.

radio items and traffic violations). Task interference is a convenient construct to investigate preferences in task prioritization, because task interference necessitates the process of task prioritization.

Threaded Cognition Theory (TCT) describes multi-tasking in terms of rapid switching (typically < 1 sec) between task goals in multiple resources (Salvucci & Taatgen, 2008). Figure 3.1 presents three main components of TCT: the goal buffer, the procedural resource, and a set of five other resources. The goal buffer holds information about the current goals of the system. Each goal ‘G’ is associated with a priority level ‘p’ (expressed in percentages) and an idle time ‘Δt’. The procedural resource selects a goal from the goal buffer when one or more other resources are available. Details on the influence of p and Δt on goal selection are described in the next paragraph. The procedural resource integrates available information from the buffers of the other resources, and initiates new goal-related behavior by sending instructions. These instructions include sampling information from the task environment (e.g., aural and visual resources), storing and retrieving information (e.g., declarative resource), and taking action in accordance with the active goal (e.g., manual and vocal resources).

TCT explains task interference through an integration of two dominant perspectives on human-information processing. In line with Wickens’ (1984; 2008) Multiple Resource Theory (MRT), task interference can take place in any of the resources. The total amount of task interference depends on the degree the demands of two tasks sharing common resources. For example, a combination of two visual/manual tasks results in more task interference than a visual/manual task with an aural/vocal task. In line with Pashler's (1994) Response Selection Bottleneck Theory (RSBT), each resource can only be used by one goal at a time. For example, the procedural resource sends an instruction to only one of the other resources at a time, and each procedural instruction requires approximately 50 ms of processing (Salvucci & Beltowska, 2008). The serial processing that results from this bottleneck causes delays when two tasks have to be performed simultaneously.

3.1.2 Goal selection in a dual-task situation

TCT literature provides two rules on goal selection by the procedural resource. First, when goals with an equal priority level simultaneously compete for the procedural resource, the least recently processed goal (i.e., with the lowest idle time Δt) claims right of way (Salvucci & Taatgen, 2008). Second, when goals have unequal priority levels, the goal with the highest priority p claims the procedural resource, whereas alternative goals have to wait until the procedural resource is available again (Salvucci & Beltowska, 2008). Furthermore, Salvucci & Beltowska suggest that a generalized view on resource scheduling can be obtained by extending the priority level from a binary variable (e.g., high vs. low) to a continuous variable. The question then becomes which of the above two rules 'wins', when multiple goals have priority levels greater than zero.

Two additional mechanisms may influence goal selection. The Memory for Goals theory (Altmann & Trafton, 2002) relates prolonged goal inactivity (i.e., several seconds) with decay in goal activation, resulting in a decreased chance of goal selection. However, we do not expect prolonged goal inactivity in concurrent dual-tasking, because goals are likely to be reselected within a few seconds (Salvucci et al., 2009). As a second mechanism, internal cues (e.g., cognitive chunking of phone numbers) and external cues (e.g., visual flow while driving) strengthen goal activation in memory (e.g., dialing, driving) (Altmann & Trafton, 2002; Janssen et al.,

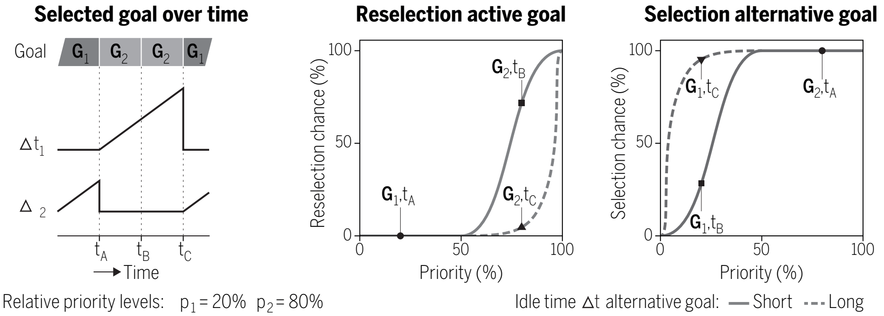

Figure 3.2. Goal selection chance. Goal selection by the procedural resource in a concurrent dual-task setting as function of goal priority p and idle time Δt since the goal was last selected.

2012). We do acknowledge that cues may influence task prioritization, but if we position these effects in the goal buffer of Figure 3.1, then describing goal selection by TCT can be confined to aforementioned rules on p and Δt.

We interpret goal selection in concurrent dual-tasking in terms of a chance mechanism. The reselection chance of an active goal G depends on its priority level p, and decreases with the idle time Δt of alternative goals. Figure 3.2 describes a dual-task scenario to illustrate the tradeoff between p and Δt. In this scenario the priority levels of each goal are fixed (i.c., $p_1 = 20\%$, $p_2 = 80\%$), whereas idle time per goal changes as function of time (see left panel). The idle time of an active goal is kept at zero (i.e., it is no longer idle when selected), whereas it increases autonomously for alternative goals (see Figure 3.1, left dashed arrow). The values of p and Δt are evaluated when the procedural resource has finished sending an instruction (e.g., at timestamps t_A, t_B, t_C in Figure 3.2). In our example, G_1 is initially the active goal. The chance of reselection at t_A equals zero, because p_1 is relatively low, and $\Delta t_2 > \Delta t_1$. Hence, G_2 becomes the active goal. At t_B the active goal G_2 is reselected, even though $\Delta t_1 > \Delta t_2$. The reason is that p_2 is relatively high. However, at t_C the S-curves have shifted in favor of the alternative goal (see dashed lines in Figure 3.2), because Δt_1 increased even further (i.e., $\Delta t_1 >> \Delta t_2$). Priority level p_2 remains relatively high, but the reselection chance drops from point 'G_2,t_B' to 'G_2,t_C'. Consequently, alternative goal G_1 has a higher selection chance, despite its low priority. This scenario illustrates how a high priority goal is reselected several times, but eventually it loses out against a low priority goal, to be selected again soon. Moreover, the scenario shows how goal priority levels can be interpreted as continuous variables within TCT, which means a preference to prioritize one task over another task does not exclude briefly attending the lower-priority task.

3.1.3 Paradigm

The experimental tasks in our study have been designed to ensure task interference, and consequently, task prioritization. Two continuous tasks have been used, based on observations in the context of police work (Jansen et al., 2014): a high speed driving task and an auditory memory task. The self-paced driving task represented police emergency driving. Participants have been given a printed map with several destinations. They had to read the map to navigate to as many destinations as possible within a fixed amount of time. The experimenter-paced memory task represented the demands of attending dispatcher-controlled police radio messages. Participants had to answer questions related to radio news items.

According to TCT, and based on empirical findings (Salvucci & Beltowska, 2008), task interference is expected to occur in shared resources. In our study, the driving task requires the visual resource (i.c., attention and processing), manual resource (i.c., motor control of the hands), declarative resource (i.c., to remember the current destination), and the procedural resource (i.c., sending instructions to the other resources). In addition, map reading requires mental rotation (Aretz & Wickens, 1992), which also places demands on the procedural resource. The memory task requires the auditory resource (i.c., attention and processing), verbal resource (i.c., to respond), declarative resource (i.c., to memorize chunks of information), and the procedural resource (i.c., to compare a memory question with the memorized chunks of information). In sum, task interference is expected in the procedural resource and the declarative resource.

Relative priority levels across the task goals determine which task suffers most from task interference (see p_1 and p_2 in Figure 3.2). However, TCT does not describe how these priority levels are set. One solution is to view task prioritization as a process on a strategic level (i.e., at a lower temporal resolution than TCT). Part of the driving skill is to strategically pay attention to other tasks for limited durations of time. Drivers can adapt the speed of their vehicle with an immediate effect on the difficulty of the (self-paced) driving task (Lansdown et al., 2004). Alternatively, they can choose to ignore the (externally paced) secondary task (Ünal et al., 2013). Our experimental setup enables such strategic leverage to control the relative priority levels between the driving task and the memory task.

In the context of police work, performance differences resulting from preferences on task prioritization should be examined over periods of time comparable with the duration of an emergency response (i.e., minutes). In this context, the millisecond time window of TCT may seem out of place. However, Salvucci & Taatgen (2008) demonstrate that TCT successfully predicts the consequences of task interference in continuous tasks, by extrapolating relatively short delays (<1 sec.) to aggregate performance measures (>>1 sec). In the present study, the presence of task interference is established by comparing aggregate performance measures of dual-task conditions with single-task conditions (O'Donnell & Eggemeier, 1986). Task preferences should be reflected in distinct tradeoffs between the proportion of destinations reached, and the proportion of correct answers. Likewise, priority instructions should result in different tradeoffs. The effectivity of preferences and priority instructions has been analyzed through the corresponding interaction effects with task conditions (i.e., dual-task versus single-task).

Experiment 1 first investigates whether task interference occurs as pre-requisite for task prioritization. This is followed by an exploration on whether participants have preferences for tasks in absence of priority instructions, and whether preferences are reflected in task performance tradeoffs. Experiment 2 replicates Experiment 1, except that the former preferences are reformulated as priority instructions, and mental effort is taken into account. Experiment 3 tests two hypotheses on why the priority instructions in Experiment 2 did not yield significant results. Finally, the findings were used to explore an integration of TCT and Hockey's (1997; 2011) Compensatory Control Model.

3.2 Experiment 1

The goal of Experiment 1 was to examine whether people have preferences in task prioritization. No task priority instructions were given. Preferences were inquired afterwards. Differences in preferences were examined by comparing the relative impact of interference between the tasks.

3.2.1 Method

Participants

Twenty-one students of the Faculty of Industrial Design Engineering volunteered (17 males, 4 females, 20 to 35 years old, average 26.1 years). This study was approved by the Ethical Committee of Delft University of Technology. Participants gave written informed consent. All were native Dutch speakers. They reported normal hearing, and normal or corrected-to-normal vision.

Auditory memory task

Twenty-seven auditory stimuli were prepared, of which three were used for training. They consisted of Dutch news items (average duration: 15.2 sec), recorded by professional newsreaders. For each news item, a factual question was recorded by a native speaker from the Netherlands. Questions were related to information items close to the center of the corresponding news item, and allowed one correct answer. For example, the item: *"In the third quarter of this year less cars were sold than in the same period of last year. To be precise: six percent less. The trade organizations also expect a decrease in sales next year."* was followed by the question: *"How many percent less cars were sold?"*. The stimuli and questions were saved as wave files (16 bit, 44.1 kHz). The goal of the memory task was to answer a question for each stimulus.

Driving task

The ‘RC Mini Racers’ (2012) game was used for the driving task. The game featured a miniature vehicle in a closed environment without moving objects. Arrow keys controlled the vehicle. A test map was created for navigation. Seventeen labelled destinations (A-Q) were added to this map, adjacent to landmarks in the driving environment (e.g., the corner of a parking lot, a billboard). In addition, a training map was created with three labelled alternative destinations. The goal of the driving task was to drive from the starting location to as many destinations as possible in alphabetical order. Each time a destination was reached, a button had to be pressed to return to the starting location. A pilot study revealed that with extensive practice, a maximum of fifteen destinations can be reached.

Apparatus

The driving game ran on an Apple MacBook Pro 15”, placed on a table in a well-lit, quiet room. The maps, printed on A3 size paper, were positioned next to the laptop. Screen activity was recorded to verify whether the car was at the correct location in each attempt. Driving sounds and auditory stimuli were played through a pair of Creative Gigaworks T20 Series II loudspeakers, positioned at ear height, and approximately 30 cm to the left and right of the laptop. The experiment was conducted using a dedicated Max program.

Measures

Auditory memory performance was calculated as the proportion of correct answers within each experimental condition. Driving performance per experimental condition was calculated as the proportion of destinations reached, where $n = 15$ corresponds with 100%. Only correct attempts were included to calculate driving performance. For example, if the vehicle was placed north of a billboard, whereas the destination on the map was south of that billboard, the attempt was evaluated as incorrect, and excluded from subsequent analysis. For statistical analysis, the proportions were transformed with an arcsine transformation (Zar, 1996). All statistical tests were conducted with SPSS v.22, and results were compared to an α level of .05. Type III sums of squares were used in all ANOVAs to compensate for differences in sample size.

Experimental design

The experiment consisted of two tasks: an auditory memory task, and a driving task. A crossover design was used with four periods, three experimental conditions, and two treatment sequences, see Fig 3.3.

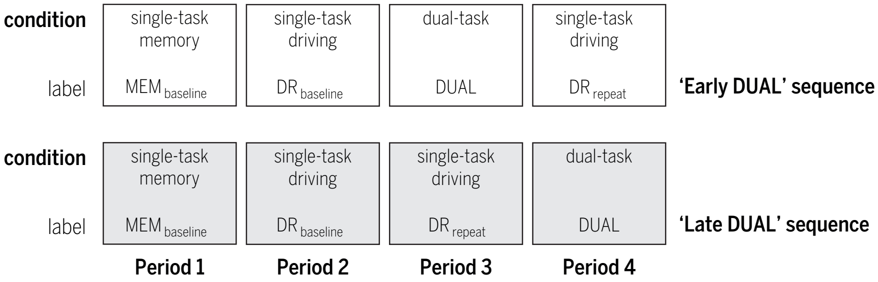

Figure 3.3. Experimental design

The first two periods concerned single-task baseline performance on the memory task (i.e., condition $\mathrm{MEM}_{\mathrm{baseline}}$), and on the driving task (i.e., condition $\mathrm{DR}_{\mathrm{baseline}}$). The remaining two periods were ordered in two sequences to discriminate between dual task effects and potential learning effects on the driving task. In the 'early DUAL' sequence, the third period was a dual-task condition (i.e., condition DUAL), and the fourth period was a repetition of the single-task driving condition (i.e., labelled as $\mathrm{DR}_{\mathrm{repeat}}$). This order was reversed in the 'late DUAL' sequence (i.e., $\mathrm{DR}_{\mathrm{repeat}}$ followed by DUAL). Participants were randomly distributed over the 'early DUAL' (n = 11) and the 'late DUAL' (n = 10) sequences. Driving task performance was analyzed by comparing Period 2, 3, and 4. Memory performance was analyzed by comparing Period 1 with the dual-task conditions in Periods 3 and 4.

Procedure

The duration of the experimental conditions (i.e., $\text{MEM}_{\text{baseline}}$, $\text{DR}_{\text{baseline}}$, $\text{DR}_{\text{repeat}}$, DUAL) was 5 minutes each. The auditory memory task ran automatically, and the driving task was self-paced. A beep sound was played to denote the end of an experimental condition.

After signing informed consent, a participant rehearsed the memory task for two minutes with three training stimuli. Memory questions were followed by a 4.5 sec answer time, a beep sound, and a 1.2 sec silence. Volumes across news items and questions were matched, and set to a comfortable listening level. A participant was instructed to verbalize an answer after each question. Responding after the beep sound was allowed if needed, but it was recommended to prepare for the next stimulus. In the $\text{MEM}_{\text{baseline}}$ condition, 12 stimuli were randomly selected per participant from 24 test stimuli, and presented in random order.

Familiarization with the driving task lasted approximately ten minutes. First, the participant drove five laps in a racing game mode to get used to the controls. Next, the navigation subtask was rehearsed on the training map, with specific attention to correct and incorrect attempts. Game sounds were included for feedback on driving speed, but their volume was set to a low level to ensure audibility of the auditory stimuli in the upcoming DUAL condition.

In the $\text{DR}_{\text{baseline}}$ condition and in subsequent conditions the training map was replaced with the test map. The execution order of the $\text{DR}_{\text{repeat}}$ and DUAL conditions depended on the allocated sequence. In the DUAL condition, the remaining 12 stimuli of the memory task were presented in random order. No task priority instructions were given. At the end of the session, a participant was asked to which task attention was mostly paid in the DUAL condition (i.e., driving task, memory task, or both), and how this allocation policy was executed.

3.2.2 Results

The presence of task interference was checked to ensure the necessity of task prioritization. Verbal reports on attention revealed two preferences regarding task prioritization. Finally, it was examined whether preferences are reflected in performance tradeoffs.

Task interference

Task interference is established when performance of one task is hindered by the addition of another task. Table 3.1 summarizes the results of a 2 (Sequence) × 2 (Period) mixed ANOVA on memory performance, and of a 2 (Sequence) × 3 (Period) mixed ANOVA on driving performance. Memory performance in the $\text{MEM}_{\text{baseline}}$ condition (i.e., Period 1) was 63.10% (SE = 3.80), see Figure 3.4A. This value indicates that the memory task was a difficult one. Memory performance dropped significantly in the DUAL condition (M = 49.60%, SE = 3.17), which implies task interference.

The car was nearby the labelled destination in 96.8% of the attempts. Only these attempts were analyzed. The maximum number of correct attempts within a period was 14, and this number was attained by one participant only. Figure 3.4B shows that driving performance increases over time. A significant main effect of Period was found. Repeated type contrasts revealed that performance increased significantly from Period 2 (M = 56.51%, SE = 3.14) to Period 3 (M = 60.95%, SE = 3.45), $F(1,19) = 5.34$, p = .032, $\eta_p^2 = .22$, as well as from Period 3 to Period 4 (M = 66.67%, SE = 3.44), $F(1,19) = 7.74$, $p < .012$, $\eta_p^2 = .30$. This finding suggests an overall learning curve on the driving task.

A significant Period × Sequence interaction demonstrates that this learning process on the driving task was negatively influenced by the presence of the auditory memory task. For this there are two indications. First, from Period 2 to 3, participants in the ‘early DUAL’ sequence show stable performance from $\text{DR}_{\text{baseline}}$ to DUAL, whereas the ‘late DUAL’ sequence shows improved performance from $\text{DR}_{\text{baseline}}$ to $\text{DR}_{\text{repeat}}$, $F(1,19) = 6.24$, p = .022, $\eta_p^2 = .25$. Second, from Period 3 to 4, the ‘early DUAL’ sequence shows

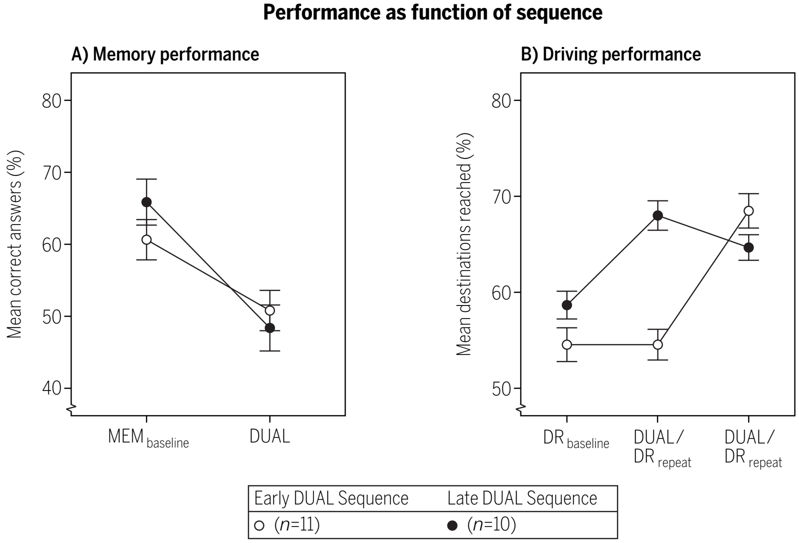

Figure 3.4. Memory task performance (A) and driving task performance (B) as function of sequence. Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability. NOTE: participants did not receive priority instructions.

Table 3.1. Summary of ANOVA results on performance as function of sequence.

	Memory performance			**Driving performance**		
Source	***F*(1,19)**	***p***	η_p^2	***F*(2,38)**	***p***	η_p^2
Period	10.57	.004	.36	13.30	< .001	.41
Sequence	.081	.78	.004	.49	.49	.025
Per × Seq	.84	.37	.042	10.52	< .001	.36

NOTE: Per = Period, Seq = Sequence.

improved performance from DUAL to $\text{DR}_{\text{repeat}}$, whereas the 'late DUAL' sequence does not from $\text{DR}_{\text{repeat}}$ to DUAL, $F(1,19) = 19.81$, $p < .001$, $\eta_p^2 = .51$. To summarize, the experimental setup resulted in bi-directional task interference. Memory performance was reduced by the addition of the driving task, whereas driving performance was hindered by the addition of the memory task.

Verbal reports on preference

Two types of verbal reports on the allocation of attention were found. Thirteen participants indicated that they paid most attention to the driving task, because they considered the driving task more rewarding, and the auditory memory task less important, and distracting. Furthermore, these participants viewed driving as an active task that could not be aborted, whereas the memory task could be ignored. We interpret these reports as a preference for the driving task (hereafter, 'driving' preference). Eight participants reported that they were motivated to perform both tasks as good as possible, and how they continuously switched attention between the tasks. We interpret these reports as an 'equal' preference for both tasks. In the 'early DUAL' sequence, the 'driving' and 'equal' preferences were found for seven and four participants, respectively. In the 'late DUAL' sequence, six participants had a 'driving' preference, and four participants had an 'equal' preference. The preference distributions were not significantly different between the 'early DUAL' and 'late DUAL' sequences ($P = 1.00$, Fisher's exact test).

Preferences versus tradeoffs

Now that two preferences regarding task prioritization have been found, the next question is whether these preferences are reflected in performance. Such reflection should be visible in the interaction between Preference and Period, because not all conditions required task prioritization. Table 3.2 summarizes the results of a 2×2×2 mixed ANOVA on memory performance, with Preference and Sequence as between-subjects factors, and Period as within-subjects factor. Table 3.2 also includes the results of a 2 (Preference) × 2 (Sequence) × 3 (Period) mixed ANOVA on driving performance. Task interference is once again demonstrated by a signifant effect of Period on memory performance, and by significant effects of Period and Period × Sequence on driving performance.

In Figure 3.5A the 'equal' preference (represented with filled circles and squares) shows stable memory performance from $\mathrm{MEM}_{\mathrm{baseline}}$ to DUAL, whereas memory performance strongly decreases with the 'driving'

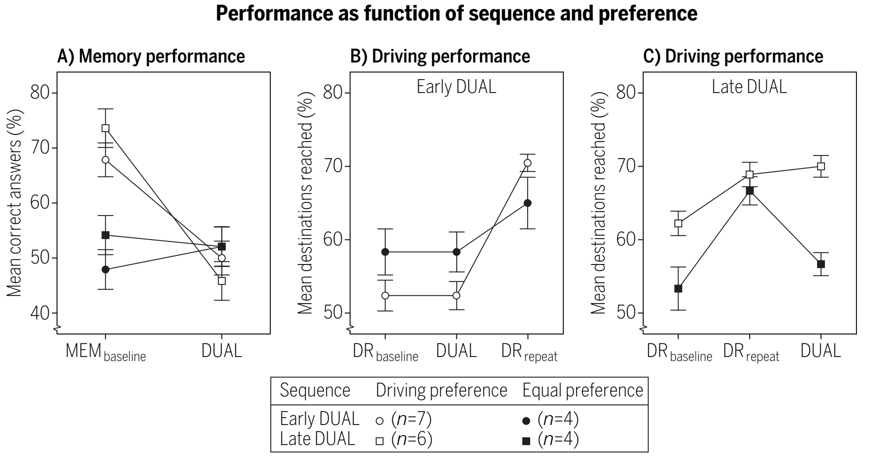

Figure 3.5. Memory task performance (A) and driving task performance (B,C) as function of sequence and preference. Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability. Note: participants did not receive priority instructions.

Table 3.2. Summary of ANOVA results on performance as function of sequence and preference.

	Memory performance			Driving performance		
Source	***F*(1,17)**	***p***	**η_p^2**	***F*(2,34)**	***p***	**η_p^2**
Period	9.45	.007	.36	12.36	< .001	.42
Sequence	.13	.72	.008	.26	.62	.015
Preference	1.88	.19	.099	.20	.66	.011
Per × Seq	1.29	.27	.070	11.47	< .001	.40
Per × Pref	11.32	.004	.40	5.64	.008	.25
Seq × Pref	.023	.88	.001	.48	.50	.027
Per × Seq × Pref	.097	.76	.006	.70	.51	.039

NOTE: Per = Period, Pref = Preference, Seq = Sequence.

preference (open circles and squares). This observation was confirmed by a significant Preference × Period interaction. In addition, participants with a 'driving' preference (*M* = 59.33%, *SE* = 3.68) appear to have a higher memory performance than those with an 'equal' preference (*M* = 51.56%, *SE* = 4.67), which is caused by differences in the MEM$_{\text{baseline}}$ condition. A separate 2 (Preference) × 2 (Sequence) ANOVA on MEM$_{\text{baseline}}$ data yielded a significant effect of Preference, $F(1,17) = 8.16$, $p = .011$, $\eta_p^2 = .32$. The other sources of variance were non-significant.

Figures 3.5B and 3.5C show the mean percentages of destinations reached for participants with an 'early DUAL' and a 'late DUAL' sequence, respectively. A significant interaction between Preference and Period was found. Repeated contrasts revealed that this interaction was only significant from Period 3 to Period 4, $F(1,17) = 13.12$, $p = .002$, $\eta_p^2 = .44$. Driving performance in the 'late DUAL' sequence shows an interaction between Preference and Period (see Figure 3.5C). Performance drops from $\mathrm{DR}_{\mathrm{repeat}}$ to DUAL for the 'equal' preference (closed squares), but not for the 'driving' preference (open squares). This interaction seems absent for participants with an 'early DUAL' sequence (see Figure 3.5B).

In summary, the presence of task interference necessitated task prioritization. Significant interactions between Preference and Period were found on both memory performance and driving performance, which demonstrates that preferences resulted in different performance tradeoffs.

3.2.3 Discussion

The two main findings of Experiment 1 are that participants have prioritization preferences in a situation of task interference (i.e., a 'driving' preference and an 'equal' preference), and that the inquired preferences are reflected in actual performance tradeoffs. The 'driving' preference inhibits task interference of the memory task on the driving task. However, this inhibition has only been found in the 'late DUAL' sequence, which suggests that increased exposure to the driving task is required for effective use of preferences.

The driving performance data strongly suggest a learning curve on the driving task, which has been accounted for by using two task sequences. Nonetheless, the learning curve may have been incomplete by the time participants performed the $\mathrm{DR}_{\mathrm{repeat}}$ ('early DUAL' sequence) or DUAL ('late DUAL' sequence) condition. As a result, it is not possible to conclude whether the 'driving' preference fully, or only partially, mitigates the interference of the memory task on the driving task. Therefore, Experiment 2 incorporates a single-task control group to investigate the learning curve on the driving task in absence of the memory task.

Perceived task utility appears to be a recurring theme in the verbal reports that were used to inquire preferences. In the transition from single-task to dual-task driving, participants with an 'equal' preference may have considered the memory task an appealing alternative to the driving task, resulting in sustained memory performance at the cost of decreased driving performance (cf. Hockey, 1997; Kurzban et al., 2013). However, it is not possible to conclude whether the ‘equal’ preference actually mitigates the interference of the driving task on the memory task, because of differences in baseline performance. One participant group may have had better memory performance skills. Another potential factor is that participants in one group have spent more effort on the task to compensate for the perceived task demands, in line with cognitive-energetic models on task performance (Hockey, 1997; Kurzban et al., 2013; Hancock & Warm, 1989; Langner & Eickhoff, 2013; Sanders, 1983). Experiments 2 and 3 address effort-related adjustment by also including measurements of mental effort.

3.3 Experiment 2

Experiment 1 revealed two preferences, which were reflected in performance tradeoffs at the late dual-task treatment sequence. The goal of Experiment 2 was to examine whether using these preferences as priority instructions results in similar performance tradeoffs. The 'late DUAL' sequence of Experiment 1 was used, because preferences were not manifested in driving performance in the 'early DUAL' sequence. A control group without any instructions was added to discriminate between dual task effects and learning effects, akin to the use of two task sequences in Experiment 1.

3.3.1 Method

The driving task and measures were identical to Experiment 1.

Participants

Thirty-four students of the Faculty of Industrial Design Engineering volunteered (25 males, 9 females, 18 to 31 years old, average 23.4 years). This study was approved by the Ethical Committee of Delft University of Technology. Participants gave written informed consent. Participants were native Dutch speakers, and reported normal or corrected-to-normal vision. No hearing problems were reported.

Auditory memory task

The number of training stimuli was increased from three to twelve to reduce potential differences in baseline performance. Apart from that, the auditory memory task was identical to Experiment 1.

Apparatus

The Max program of Experiment 1 was extended with the subjective Rating Scale Mental Effort (RSME) (Zijlstra, 1993). This scale has a range from 0 to 150, and is accompanied by Dutch anchor words.

Experimental design

Participants were randomly distributed over a 'driving' instruction (*n* = 12), an 'equal' instruction (*n* = 11), and a control group without an instruction (*n* = 11). The 'late DUAL' sequence of Experiment 1 was used for the 'driving' and 'equal' instruction groups: $\mathrm{MEM}_{\mathrm{baseline}}$-$\mathrm{DR}_{\mathrm{baseline}}$-$\mathrm{DR}_{\mathrm{repeat}}$-$\mathrm{DUAL}_{\mathrm{instr}}$. The control group did not include a dual-task condition, but instead it featured two additional single-task conditions: $\mathrm{MEM}_{\mathrm{baseline}}$-$\mathrm{DR}_{\mathrm{baseline}}$-$\mathrm{DR}_{\mathrm{repeat}}$-$\mathrm{MEM}_{\mathrm{repeat}}$-$\mathrm{DR}_{\mathrm{repeat2}}$.

Procedure

Two modifications were made to the procedure of Experiment 1. A priority instruction was given before the $\mathrm{DUAL}_{\mathrm{instr}}$ condition. Participants with the ‘driving’ instruction had to prioritize the driving task. They were invited to perform the memory task, but only if this would not degrade driving task performance. Participants with an ‘equal’ instruction had to treat both tasks as equally important by performing as good as possible on both tasks. Participants in the control group did not receive a priority instruction, because no dual-task condition was involved. Finally, subjective mental effort was administered after each condition with an onscreen RSME.

3.3.2 Results

Analogous to Experiment 1, the presence of task interference was checked to ensure the necessity of task prioritization. This was followed by an examination into the effect of the 'driving' and 'equal' instructions on tradeoffs between performance and mental effort. Finally, the control group of this experiment was compared with the data of Experiment 1 to investigate learning effects.

Task interference

Figures 3.6A and 3.6B show memory performance and driving performance, respectively. For driving performance, the maximum number of correct attempts within a period was 15, and only these correct attempts were analyzed. Figures 3.6C and 3.6D show mental effort related to the memory task and the driving task, respectively. Across these graphs the same mental effort data are used for the 'driving' and 'equal' instructions in the $\text{DUAL}_{\text{instr}}$ condition. For the control group, however, the mental effort data of the single-task $\text{MEM}_{\text{repeat}}$ and $\text{DR}_{\text{repeat2}}$ conditions are used for comparisons in the $\text{DUAL}_{\text{instr}}$ condition.

Figure 3.6A shows that, for the 'driving' and 'equal' instructions, memory performance clearly decreases from $\text{MEM}_{\text{baseline}}$ to $\text{DUAL}_{\text{instr}}$, whereas the control group shows stable performance. Driving performance in Figure 3.6B increases similarly from $\text{DR}_{\text{baseline}}$ to $\text{DR}_{\text{repeat}}$ for all groups, and then remains relatively stable from $\text{DR}_{\text{repeat}}$ to $\text{DUAL}_{\text{instr}}$, whereas the performance tends to increase for the control group. For the 'driving' and 'equal' instructions, these transitions come at the expense of increased

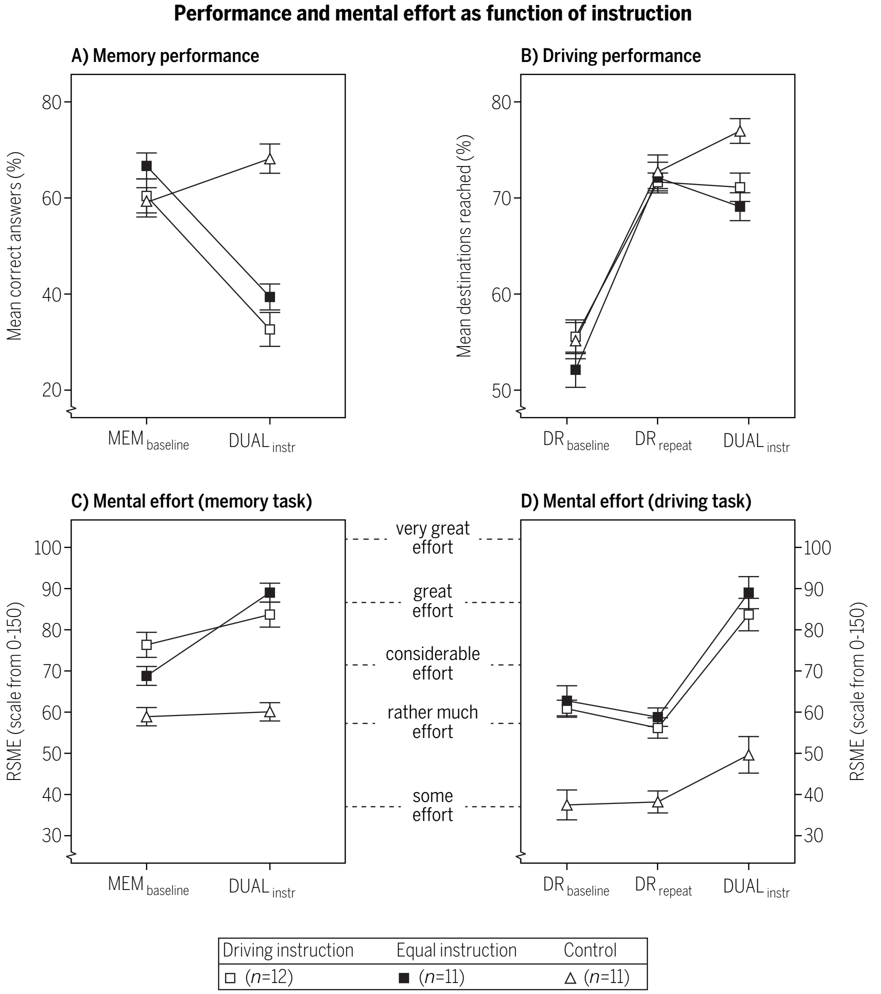

Figure 3.6. Memory task performance (A), driving task performance (B) and subjective mental effort (C,D) as function of instruction. Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subject variability.

mental effort from $\text{MEM}_{\text{baseline}}$ (anchor word: 'considerable effort') to $\text{DUAL}_{\text{instr}}$ (anchor word: ‘great effort’), and increased mental effort from $\text{DR}_{\text{repeat}}$ (anchor word: ‘rather much effort’) to $\text{DUAL}_{\text{instr}}$ (anchor word: 'great effort') (see Figures 3.6C and 3.6D). The control group, however, shows stable mental effort on the memory task, and relatively stable mental effort on the driving task. At both tasks the ratings of the control group appear to be lower than the other instruction groups, with substantial higher mental effort for the memory task than for the driving task. The observed tradeoff between memory performance and mental effort indicates dual-task interference between the memory task and the driving task.

These observations were supported by the results of a mixed 3 (Instruction) × 2 (Period) ANOVA on memory performance, and a mixed 3 (Instruction) × 3 (Period) ANOVA on driving performance. Both ANOVAs were also conducted on mental effort, corresponding with the memory task (i.e., 2 periods) and the driving task (i.e., 3 periods). Table 3.3 summarizes the results of these tests.

Memory performance decreased significantly from $\text{MEM}_{\text{baseline}}$ to $\text{DUAL}_{\text{instr}}$, but a significant interaction between Period and Instruction shows that this was not the case for the control group. The above interaction was also significant on mental effort with the memory task. Figure 3.6C suggests that mental effort increases with the 'equal' instruction, whereas it remains stable in the control group.

Table 3.3. Summary of ANOVA results on performance and mental effort as function of instruction.

	Memory performance				**Driving performance**			
Source	*df*	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2
Period	(1,31)	17.28	< .001	.36	(2,62)	79.44	< .001	.72
Instruction	(2,31)	3.02	.064	.16	(4,62)	.31	.74	.019
Per × Instr	(2,31)	11.71	< .001	.43	(4,62)	1.89	.12	.11
	Mental effort (memory task)				**Mental effort (driving task)**			
	df	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2
Period	(1,31)	3.56	.068	.10	(1.51,46.77)	29.14	< .001	.49
Instruction	(2,31)	5.29	.011	.25	(2,31)	10.63	< .001	.41
Per × Instr	(2,31)	3.80	.033	.20	(3.02,46.77)	1.65	.17	.096

NOTE: Instr = Instruction, Per = Period. The *df* of mental effort on the driving task were adjusted using Greenhouse-Geisser, $\varepsilon = .75$.

Significant main effects of Period were found on driving performance, and on mental effort with the driving task. Repeated contrasts showed that driving performance increased significantly from $DR_{baseline}$ to DR_{repeat}, $F(1,31) = 111.88$, $p < .001$, $\eta_p^2 = .78$, but not from DR_{repeat} to $DUAL_{instr}$, *n.s.* Mental effort, on the other hand, only increased significantly from DR_{repeat} to $DUAL_{instr}$, $F(1,31) = 44.51$, $p < .001$, $\eta_p^2 = .59$. Figure 3.6D indicates that the 'driving' and 'equal' instruction groups were the main drivers for this effect. Thus, participants improved their driving performance without investing more mental effort, but mental effort increased when the memory task was added. The interaction between Period and Instruction on driving performance was non-significant, which indicates that the 'driving' and 'equal' instruction groups followed a similar learning curve as the control group.

In addition, two significant main effects of Instruction on mental effort were found. Figures 3.6C and 3.6D show that for both tasks the mental effort ratings of the control group are lower than the other instruction groups. Furthermore, Figures 3.6C and 3.6D suggest that mental effort was higher in the MEMbaseline condition than in the $DR_{baseline}$ condition. This difference was confirmed through a two-way ANOVA with Instruction and Task as factors, which yielded a significant effect on Task, $F(1,31) = 21.08$, $p < .001$, $\eta_p^2 = .41$. This finding suggests that the memory task placed a heavier burden in the $DUAL_{instr}$ condition than the driving task.

Instructions versus tradeoffs

Figure 3.6 shows a high degree of similarity on all measures between the ‘equal’ and ‘driving’ priority instructions. Although significant interactions between Instruction and Period were found, these were all related to differences with the control group. This also applies to the significant main effects of Instruction on mental effort. The absence of significant differences between the 'driving' and 'equal' instructions was not caused by differences between participant groups, as they performed similar in the single-task conditions, and showed similar mental effort ratings.

Comparison with learning curves from Experiment 1

The control group of Experiment 2 helps to understand the apparent learning curves in Experiment 1. Within the control group, a *t*-test did not reveal a significant difference in memory performance between MEM$_{\text{baseline}}$ (*M* = 59.09%, *SE* = 4.42) and MEM$_{\text{repeat}}$ (*M* = 68.18%, *SE* = 5.60). Furthermore, a repeated-measures one-way ANOVA showed a significant effect of Period on driving performance, $F(2,20) = 26.67$, $p < .001$, $\eta_p^2 = .73$. Driving performance increased significantly from DR$_{\text{baseline}}$ (*M* = 55.15%, *SE* = 4.41) to DR$_{\text{repeat}}$ (*M* = 72.73%, *SE* = 3.42), $F(1,10) = 25.89$, $p < .001$, $\eta_p^2 = .72$, but not from DR$_{\text{repeat}}$ to DR$_{\text{repeat2}}$ (*M* = 76.97%, *SE* = 4.25), *n.s.* These findings suggest that there is no learning curve on the memory task, whereas two experimental periods are required to fully learn the driving task.

In Figure 3.7 the control group is juxtaposed with the 'driving' and 'equal' preferences in the 'late DUAL' sequence of Experiment 1. Figure 3.7A shows that memory performance decreases with the 'driving' preference, whereas it remains relatively stable with the 'equal' preference and in the control group. All groups appear to have reached a similar driving performance level in the DR$_{\text{repeat}}$ condition (see Figure 3.7B), which is consistent with the above statement on the driving task learning curve. Furthermore, driving performance decreases strongly with the 'equal' preference from DR$_{\text{repeat}}$ to DUAL$_{\text{instr}}$, whereas it remains stable with both the 'driving' preference and the control group.

The results of a 3×2 mixed ANOVA on memory performance and a 3×3 mixed ANOVA on driving performance support these observations (see Table 3.4). On memory performance a significant interaction between Preference and Period was found. In addition, a one-way ANOVA on the MEM$_{\text{baseline}}$ condition did not reveal a significant difference in baseline performance between the preferences and the control group. This implies that participants with the 'equal' preference managed to protect memory performance as if no additional task was involved.

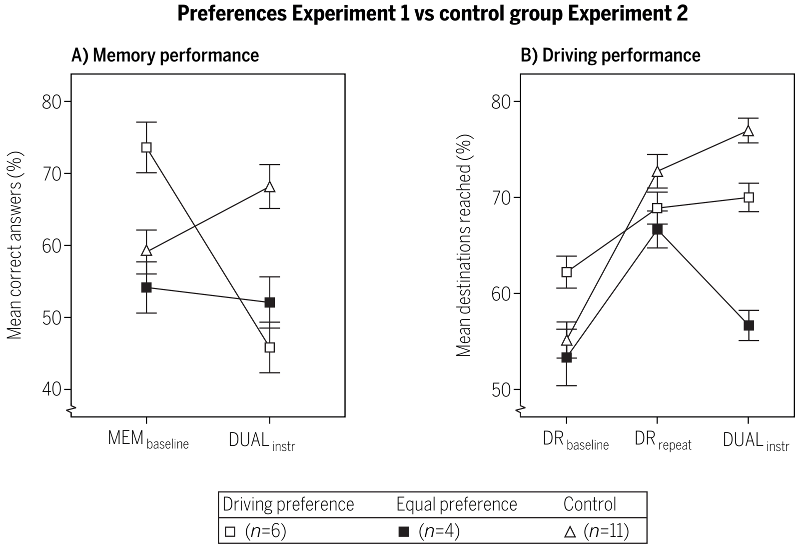

Figure 3.7. Performance by the control group (Experiment 2) versus two preferences (Experiment 1). Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability. NOTE: participants did not receive priority instructions.

Table 3.4. Summary of ANOVA results on the control group (Experiment 2) versus two preferences (Experiment 1).

	Memory performance				Driving performance			
Source	*df*	*F*	*p*	η_p^2	*df*	*F*	*p*	η_p^2
Period	(1,18)	1.93	.18	.097	(2,36)	19.02	< .001	.51
Preference	(2,18)	.87	.44	.088	(2,18)	.94	.41	.095
Per × Pref	(2,18)	7.36	.005	.45	(4,36)	5.15	.002	.36

NOTE: Per = Period, Pref = Preference.

The overall learning curve on driving performance was reflected in a significant main effect of Period. This effect was significant from DR$_{\text{baseline}}$ to DR$_{\text{repeat}}$, $F(1,18) = 27.15$, $p < .001$, $\eta_p^2 = .60$, but not from DR$_{\text{repeat}}$ to DUAL$_{\text{instr}}$, *n.s.* The interaction between Preference and Period was significant only from DR$_{\text{repeat}}$ to DUAL$_{\text{instr}}$, $F(2,18) = 5.04$, $p = .018$, $\eta_p^2 = .36$. A separate one-way ANOVA on the DR$_{\text{repeat}}$ condition did not yield a significant effect, which means that the preference groups learned to perform the driving task at a similar level as the control group.

3.3.3 Discussion

Experiment 2 has two main findings: a substantial increase in mental effort from single-task to dual-task conditions, and no effect of the manipulation of the 'driving' and 'equal' priority instructions. A comparison with the late dual-task sequence of Experiment 1 clarifies which instruction has not been followed. The ‘driving instruction’ shows a performance tradeoff similar to the ‘driving’ preference group: stabilized driving performance at the expense of decreased memory performance. A comparison with the control group confirms that decreased memory performance in both instruction groups could be attributed to dual-task interference.

Contrary to the 'driving' instruction group, the ‘equal’ instruction group shows a performance tradeoff dissimilar to its ‘equal’ preference counterpart. In fact, it resembles the performance tradeoff of the ‘driving’ preference group. Therefore, participants in Experiment 2 appear to have followed the ‘driving’ instruction, but not the ‘equal’ priority instruction. This is in line with the observed 3:2 distribution of the ‘driving’ and ‘equal’ preferences in Experiment 1, suggesting a majority of the participants in the ‘equal’ instruction group prefer the ‘driving’ instruction, and acting accordingly.

Next to these performance tradeoffs, it appears there has also been a tradeoff between performance and mental effort. In the transition from single-task to dual-task conditions, driving performance remains stable, but at the cost of decreased memory performance and increased mental effort. This tradeoff can be interpreted as a protection mechanism of the driving task against performance degradation. Such a protection mechanism has been described previously by the Compensatory Control Model (1997; 2011), which predicts strategies involving secondary task decrements and increased mental effort. Interestingly, participants with a ‘driving’ preference in Experiment 1 reported the memory task as secondary to the driving task.

An additional factor that may explain why the 'equal' instruction was not followed is related to the experimental design. Potential effects resulting from the priority instructions may have been overshadowed by the increased demands associated with the single-task to dual-task transition. Support is found in a study by Liepelt et al. (2011) on the effect of dual-task exposure on intertask coordination. The researchers let one participant group train two tasks separately (e.g., a visual/manual and an auditory/vocal task), whereas another group received a mixture of single-task and dual-task training conditions. Participants were instructed to prioritize both tasks equally. The latter group outperformed the former group on the auditory/vocal task in a dual-task test condition. Improved dual-task performance was related to accelerated task switching in the response selection stage (cf. RSBT), which could only be trained during dual-task conditions. These findings suggest that the 'equal' priority instruction in the present study may be effective after additional dual-task exposure, especially in relation to auditory memory performance.

3.4 Experiment 3

The goal of Experiment 3 was to juxtapose preferences with priority instructions in the same experimental setup. Like the previous experiment, the manipulation of the priority instructions was evaluated through the interaction between Instruction and Period. However, this time two dual-task conditions were used: one condition without priority instructions, and one condition with. The possibility of conflicting preferences was taken into account by asking participants afterwards about their preference in the first dual-task condition.

In Experiment 2 we compared task performance with Experiment 1 to evaluate the resemblance between the priority instructions and the preferences on which the priority instructions were based. The addition of a second dual-task condition in Experiment 3 no longer allows for such a comparison with Experiment 1. Therefore, a 'free choice' group was added, that will not receive a priority instruction during the second dual-task condition.

3.4.1 Method

The auditory memory task, driving task, apparatus, and measures, were identical to Experiment 2.

Participants

Forty-three students of the Faculty of Industrial Design Engineering volunteered for a €10,- reward (29 males, 14 females, 18 to 28 years old, average 21.3 years). This study was approved by the Ethical Committee of Delft University of Technology. Participants gave written informed consent. All were native Dutch speakers. They reported normal hearing, and normal or corrected-to-normal vision.

Experimental design

Participants were randomly distributed over three priority instructions: 'driving' (n = 14), 'equal' (n = 15), or 'free choice' (e.g., no instruction at all, n = 14). The following sequence was used: $\mathrm{DR}_{\mathrm{baseline}}$-$\mathrm{DUAL}_{\mathrm{baseline}}$-$\mathrm{DUAL}_{\mathrm{instr}}$, in which $\mathrm{DUAL}_{\mathrm{baseline}}$ concerned a dual-task baseline condition. The $\mathrm{MEM}_{\mathrm{baseline}}$ condition was removed to ensure equal exposure across all Experiments. Such removal is legitimate, because the control group in Experiment 2 showed stable performance on the memory task.

Procedure

The procedure of Experiment 2 was modified. The memory task and the driving task were practiced as before (i.e., 12 training stimuli, a separate training map with 3 destinations). No priority instructions were given, except in the $\mathrm{DUAL}_{\mathrm{instr}}$ condition. At the end of the session, the participant was asked to which task attention was mostly paid in the $\mathrm{DUAL}_{\mathrm{baseline}}$ condition (i.e., driving task, memory task, or both tasks), and how this was executed.

3.4.2 Results

One participant with the 'equal' instruction and one participant in the 'free choice' group were excluded from analysis, because they were unable to execute the tasks. First, we examined how the priority instructions were followed. Subsequent analyses investigated whether preferences influenced how these instructions were followed.

Instructions versus tradeoffs

A 3 (Instruction) × 2 (Period) mixed ANOVA was conducted to investigate the influence of Instruction on tradeoffs between performance and mental effort, see Table 3.5. Driving performance increased significantly from $\text{DUAL}_{\text{baseline}}$ to $\text{DUAL}_{\text{instr}}$, which is indicative for a learning effect (see Figure 3.8B). Furthermore, the interaction between Instruction and Period proved to be significant on all measures. Figure 3.8 shows that memory performance increases with the 'equal' instruction, at the cost of increased mental effort, and with stable driving performance. The 'driving' instruction, on the other hand, shows increased driving performance and decreased mental effort, at the cost of slightly decreasing memory performance. Finally, the 'free choice' group appears to mirror the 'driving' group on memory performance, but the 'equal' group on mental effort. The tradeoffs between memory performance and driving performance with the 'driving' and 'equal' instructions are in line with those found in Experiment 1, which means the instructions were followed as intended.

Verbal reports on preference

Although the instructions were apparently followed, participants may have differed in their preferences regarding task prioritization within each instruction group. The verbal reports of twenty-six participants on the $\text{DUAL}_{\text{baseline}}$ condition were interpreted as ‘driving’ preference. These participants noted that the driving environment provided stronger cues than the news items in the background, that the implications of not paying attention to the driving task were more immediate, and that standing still

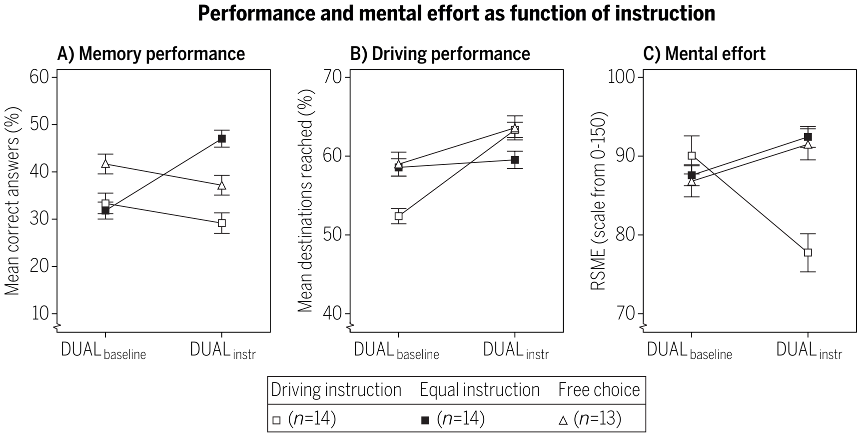

Figure 3.8. Performance and mental effort as function of instruction. Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability. NOTE: no instruction was provided in the DUAL$_{\text{baseline}}$ condition.

Table 3.5. Summary of ANOVA results on performance and mental effort as function of instruction.

		Memory performance			**Driving performance**			**Mental effort**		
Source	***df***	***F***	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Period	(1,38)	.82	.37	.021	15.23	< .001	.29	.18	.68	.005
Preference	(2,38)	1.31	.28	.065	.18	.84	.009	.45	.64	.023
Per × Pref	(2,38)	7.02	.003	.27	3.93	.028	.17	6.52	.004	.26

NOTE: Per = Period, Pref = Preference.

was not an option. The driving task was also prioritized because it was considered easier and more interesting, whereas the news items were considered irrelevant during driving.

Fourteen verbal reports were interpreted as 'equal' preference. These participants reported a desire to combine the two tasks, and to avoid incorrect answers while reaching as many destinations as possible. Their approaches were described as driving slower to perform both tasks at the same time, and to frequently switch attention, but it was also noted that attending news items occasionally resulted in losing track on the driving task.

In addition, one participant in the 'free choice' group appeared to prefer the memory task. This participant showed results comparable to the 'driving' and 'equal' preference groups within the 'free choice' instruction, except that memory performance was relatively high (i.c., 67% at $\text{DUAL}_{\text{baseline}}$, 79% at $\text{DUAL}_{\text{instr}}$). Although a preference for the memory task apparently exists, its occurrence is rare (also see Experiments 1 and 2). Therefore, further analysis is restricted to the 'driving' and 'equal' preferences.

Table 3.6 shows the resulting distribution of preferences. No significant differences were found in the preference distributions between the instruction groups (*P* = .78, Fisher's exact test). In addition, the preference distribution within the 'free choice' group was not significantly different from the preference distribution in Experiment 1 (*P* = 1.00, Fisher's exact test). The next question, then, is whether these preferences influenced how the instructions were followed, just as they affected performance tradeoffs in Experiment 1.

Table 3.6. Participant distribution as function of task priority instruction and preference.

Priority instruction	Preference: driving	Preference: equal	Total
Driving	8	6	14
Equal	10	4	14
Free choice	8	4	12
Total	26	14	40

NOTE: Participants in the free choice group did not receive a task priority instruction. Not reported in this table is one participant in the free choice group, who preferred to prioritize the memory task.

Instructions versus preferences

The 'driving' and 'equal' instruction groups are compared to whether preferences influence how the instructions are followed. The 'free choice' group is omitted from this comparison, because conflicts with preferences are not applicable without an instruction. Figure 3.9 displays task performance and mental effort as function of Instruction and Preference.

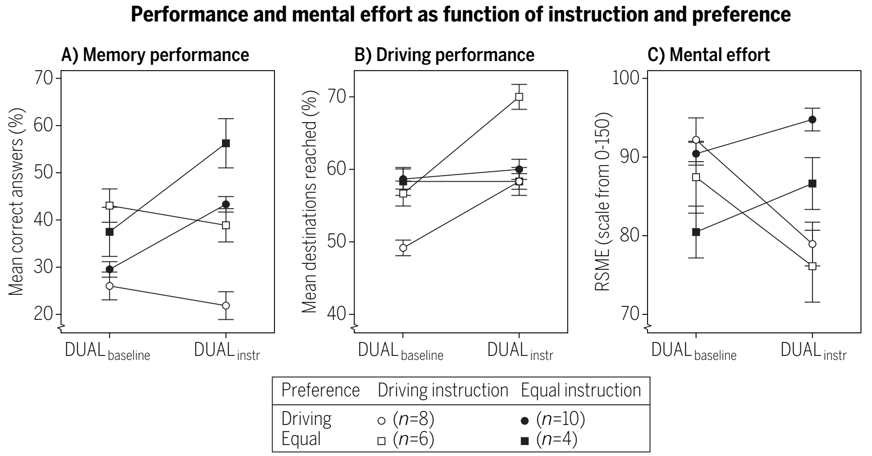

Figure 3.9. Results of the ‘driving’ and ‘equal’ priority instruction groups as function of preference. Lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subject variability. The participant with a preference for the memory task in the 'free choice' group was omitted. NOTE: no instruction was provided in the DUAL$_{\text{baseline}}$ condition.

Table 3.7. Summary of ANOVA results on performance and mental effort as function of instruction and preference.

	Memory performance			**Driving performance**			**Mental effort**		
Source	***F***	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Period	3.32	.081	.12	14.25	.001	.37	1.34	.26	.053
Instruction	2.51	.13	.095	.005	.95	< .001	.52	.48	.021
Preference	5.85	.024	.20	.42	.52	.017	1.10	.31	.044
Per × Instr	10.06	.004	.30	9.83	.004	.29	8.35	.008	.26
Per × Pref	.31	.59	.013	.077	.78	.003	.096	.76	.004
Instr × Pref	.30	.59	.012	.69	.41	.028	.18	.67	.008
Per × Pref × Instr	.13	.73	.005	.69	.42	.028	< .001	.99	< .001

NOTE: *df* = (1,24). Instr = Instruction, Per = Period, Pref = Preference.

As before, the priority instructions clearly caused different tradeoffs between performance and mental effort. The ‘equal’ instruction (closed symbols) shows increasing memory performance and stable driving performance, at the expense of increasing mental effort. By contrast, the ‘driving’ instruction (open symbols) shows slightly decreasing memory performance, increasing driving performance, and decreasing mental effort. A 2 (Instruction) × 2 (Preference) × 2 (Period) mixed ANOVA confirmed these observations, with a significant interaction between Instruction and Period on all measures (see Table 3.7).

Within each instruction group, all increments and decrements are in the same direction for both preferences (i.e., comparing circles vs. squares per instruction). As a result, no significant Preference × Period interactions were found, nor were there significant Instruction × Preference × Period interactions. The significant Instruction × Period interactions suggest that participants were able to follow the task priority instructions. Moreover, the absence of other significant interactions implies that task priority instructions were followed, regardless of preference.

Nonetheless, the magnitude with which the preferences separate within each instruction group (i.e., compare Figures 3.8 and 3.9) indicates that preferences did affect absolute performance and mental effort. In Figure 3.9A memory performance is higher with the 'equal' preference than with the 'driving' preference in both instruction groups. This was supported by a significant main effect of Preference on memory performance. No other main effects of Preference were found.

A significant effect of Period was found on driving performance. Figure 3.9B shows that the main driver for this effect is the 'driving' instruction group. Note, however, that the absolute performance level in all groups is still below that of the control group in Experiment 2 (see Figure 3.6B). If the task is fully learned, then the 'driving' instruction is expected to result in stable driving performance, whereas a decrement is expected with the 'equal' instruction (see Figure 3.7B). Therefore, the main effect of Period in the present experiment can be interpreted as a learning curve.

A closer inspection of the $\mathrm{DUAL}_{\mathrm{baseline}}$ condition in Figure 3.9 indicates that the various groups differ in their baseline performance and mental effort. For example, in Figure 3.9B the group with a 'driving' instruction and a 'driving' preference shows lower driving performance in the $\mathrm{DUAL}_{\mathrm{baseline}}$

condition than the other groups. This suggests that participants were not sufficiently trained to reach an equal performance level before being exposed to the dual-task conditions. We tested this observation by subjecting the $\mathrm{DUAL}_{\mathrm{baseline}}$ data to a one-way ANOVA with four levels (i.e., the logical combinations of Instruction and Preference). In addition, a one-way ANOVA with four levels was conducted on $\mathrm{DR}_{\mathrm{baseline}}$ and on the memory training data to examine single-task differences. No significant effects were found in either test. It seems that participants were not yet fully trained on the driving task before the $\mathrm{DUAL}_{\mathrm{instr}}$ condition, but they were equally trained across the groups.

Resemblance between instructions and preferences

The previous section compared the ‘driving’ and ‘equal’ instruction groups to demonstrate that priority instructions were followed regardless of preferences. This section also includes the ‘free choice’ group, to investigate whether priority instructions resulted in task performance and mental effort comparable with the preferences on which the instructions were based. Within the ‘free choice’ group itself, participants with a ‘driving’ preference had lower memory performance and higher driving performance on the $\mathrm{DUAL}_{\mathrm{baseline}}$ and $\mathrm{DUAL}_{\mathrm{instr}}$ conditions. However, a 2 (Preference) × 2 (Period) ANOVA yielded no significant effects for both measures. Similarly, no significant effects were found on mental effort.

Two separate 2 (Instruction) × 2 (Period) mixed ANOVAs were conducted. One ANOVA concerned participants with a 'driving' preference within the 'driving' and 'free choice' instruction groups. The other ANOVA concerned participants with an 'equal' preference within the 'equal' and 'free choice' instruction groups. The results of these tests are summarized in Table 3.8. Participants with a 'driving' preference showed signicantly higher performance in the $\mathrm{DUAL}_{\mathrm{instr}}$ condition ($M = 61.25\%$, $SE = 4.64$) than in the $\mathrm{DUAL}_{\mathrm{baseline}}$ condition ($M = 56.25\%$, $SE = 4.55$). This effect reflects the learning curve on the driving task. Furthermore, a significant interaction between Instruction and Period was found on mental effort, again for participants with a 'driving' preference. The 'driving' instruction resulted in

Table 3.8. Summary of ANOVA results on preferences with matching instructions.

	Preference: 'driving'								
	Memory performance			**Driving performance**			**Mental effort**		
Source	***F***	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Period	1.17	.30	.077	4.88	.044	.26	.29	.60	.020
Instruction	2.16	.16	.13	1.22	.29	.080	.54	.48	.037
Per × Instr	.005	.94	< .001	3.47	.084	.20	7.55	.016	.35
	Preference: 'driving'								
	Memory performance			**Driving performance**			**Mental effort**		
	F	***p***	η_p^2	***F***	***p***	η_p^2	***F***	***p***	η_p^2
Period	2.01	.21	.25	2.03	.20	.25	.92	.37	.13
Instruction	.001	.97	< .001	1.91	.22	.24	.005	.95	.001
Per × Instr	1.99	.21	.25	.15	.71	.025	.72	.43	.11

NOTE: Participants with an 'equal' instruction were excluded from the ANOVA on the 'driving' preference. Vice versa, participants with a 'driving' preference were excluded from the ANOVA on the 'equal' preference. *df* for preference 'driving': (1,14). *df* for prefrence 'equal': (1,6). Instr = Instruction, Per = Period.

decreased mental effort from $\text{DUAL}_{\text{baseline}}$ (*M* = 92.19, *SE* = 5.02) to $\text{DUAL}_{\text{instr}}$ (*M* = 78.94, *SE* = 6.76), whereas the 'free choice' group showed increased mental effort from $\text{DUAL}_{\text{baseline}}$ (*M* = 88.98, *SE* = 10.52) to $\text{DUAL}_{\text{instr}}$ (*M* = 97.88, *SE* = 8.99). No other significant effects were found.

To summarize, participants who acted according to their preference showed task performance similar to those with a matching instruction. The 'driving' instruction, however, resulted in decreased mental effort compared to the 'free choice' group. The latter group may have had doubts on how well they were expected to perform on the memory task. The presence of the 'driving' instruction may have resulted in more efficient use of energetic resources.

3.4.3 Discussion

Experiment 3 has three main findings. Priority instructions have been followed, regardless of preference. Nonetheless, preference does influence memory performance, regardless of the instruction. Finally, the instructions have resulted in performance that resembles the preferences on which the instructions have been based. These findings lead back to the question why the 'equal' priority instruction was not followed in Experiment 2. We formulated two explanatory factors: conflicting preferences and lack of dual-task exposure. The successful manipulation of priority instructions in Experiment 3 appears to refute the factor of conflicting preferences.

However, through logical reasoning it must be concluded that both factors play a role. Suppose that preference has no effect on task performance. In that case, a lack of dual-task exposure would be the only explanation why participants in Experiment 2 have been unable to follow the priority instructions. However, the same amount of dual-task exposure has been given to participants in Experiment 1, yet they have been able to act according to their preference. This means preference must have played a role in Experiment 2.

Now suppose that preference is the only factor that has influenced following priority instructions in Experiment 2. In that case, an equally disruptive effect of preference would be expected in Experiment 3. Although the 'equal' preference has shown improved memory performance, also in the 'driving' instruction group, its influence has been too small to hinder the priority instructions. This means preference cannot be the only factor that influences following instructions. Together with the previous deduction, this suggests that the increased amount of dual-task exposure in Experiment 3 has decreased the effect of conflicting preferences on following priority instructions.

3.5 General discussion

The central question in this study was whether people differ in their preferences regarding task prioritization, and if so, whether these preferences influence the effectiveness of priority instructions. The results of three experiments show that people indeed have distinct preferences in an experimental dual-task setting (Experiment 1), which can be overruled

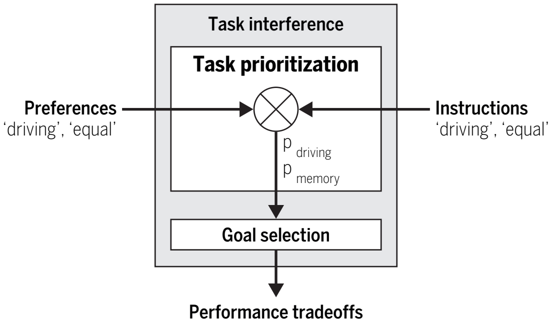

Figure 3.10. Model of task prioritization in the context of task interference.

by priority instructions, but only after a certain amount of dual-task exposure (Experiments 2 and 3).

Figure 3.10 provides an overview of the phenomena in this study. Combining two tasks with overlapping resources has created a situation of task interference. Performance tradeoffs are a direct consequence of task interference, in that task performance on one or both tasks is lower compared to single-task performance. A task prioritization process regulates which of the tasks suffers most from task interference, by setting priority levels for each task goal (Gopher & Navon, 1980; Norman & Bobrow, 1975). These priority levels in turn influence goal selection by the procedural resource, as illustrated previously in Figure 3.2. Experiment 1 demonstrates that preferences (i.c., 'driving', 'equal') influence the task prioritization process (i.e., the levels of pdriving and pmem), because these preferences have resulted in distinct performance tradeoffs. By contrast, the task prioritization processes in Experiments 2 and 3 have not only been a function of intrinsic preferences, but also of extrinsic instructions. We thus observed that the 'equal' instruction was not followed in Experiment 2, but it was followed in Experiment 3, after increased dual-task exposure. From this we speculated that if both preferences and instructions influence task prioritization, the relative weights of these factors on the priority levels should determine whether tasks are performed in favor of the instruction, or the preference. The next section summarizes how the weights of preferences and instructions on the relative task priority levels have differed between the experiments. This gives rise to an integration of TCT’s goal selection mechanism within a framework of regulatory control.

3.5.1 Variable weight of preferences

Preference appears to have affected the relative priority levels of each task goal (hereafter, ‘priority distribution’) with different weights throughout the experiments. Figure 3.11 shows an hypothetical priority distribution between the driving task and the memory task for each experiment, to illustrate our speculation on relative differences across the instructions and experiments.

In Experiment 1 preference has been responsible for distinct priority distributions (see arrow ‘1’). We interpret the ‘equal’ preference as a 50/50% distribution between the driving task (i.e., gray bars in Figure 3.11) and the memory task (i.e., white bars). The ‘driving’ preference cannot be represented as a 100/0% distribution, because memory performance scores above zero demonstrate that the memory task was still attended. Therefore,

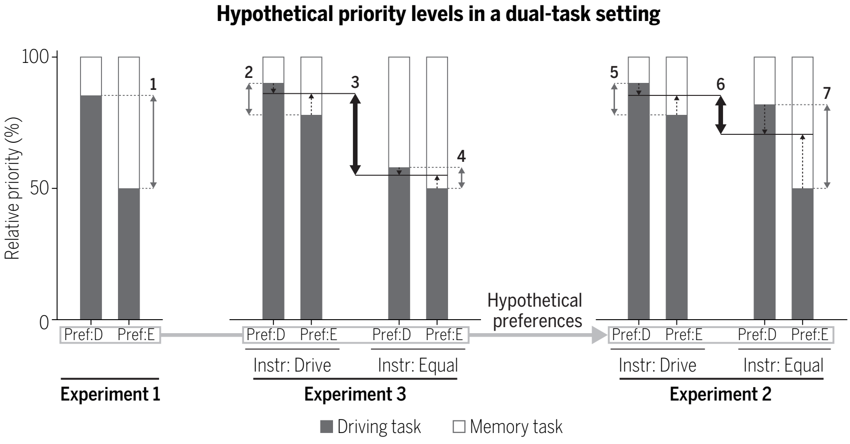

Figure 3.11. Hypothetical priority levels in three experiments. Pref:D and Pref:E correspond with ‘driving’ and ‘equal’ preferences, respectively. The preference distribution in Experiment 2 is based on Experiments 1 and 3. Arrows (1,2,4,5,7) correspond with differences in priority allocation as result of preference. Arrows (3,6) indicate differences as result of priority instruction, averaged over the number of preferences within each instruction. Dashed arrows point to the weighted average of preferences within an instruction. Numbered arrows are described in the text.

we interpret the ‘driving’ preference as an 80/20% priority distribution in favor of the driving task, corresponding with the priority levels p_2 and p_1 in the example of Figure 3.2.

An instruction should result in a similar priority distribution as the preference on which the instruction is based. In Experiment 3 differences in priority distribution have been caused by priority instructions (arrow ‘3’). Nonetheless, the higher memory performance with the ‘equal’ instruction demonstrates that preference did influence priority distribution (arrows ‘2’ and ‘4’). Therefore, it is safe to assume that preference has also played a role in Experiment 2 (arrows ‘5’ and ‘7’). In addition, the consistent distribution of preferences in Experiments 1 and 3 suggests that in Experiment 2, too, the majority of participants has had a ‘driving’ preference. These assumptions explain why the ‘driving’ instruction in Experiment 2 has resulted in a similar performance tradeoff as the ‘driving’ preference in Experiment 1. Moreover, if the majority of participants with an ‘equal’ instruction have acted according to their ‘driving’ preference, it becomes clear why task performance and mental effort did not deviate significantly from the ‘driving’ instruction (i.e., arrow ‘6’ is small compared to arrow ‘3’).

The variable weight of preferences may be explained by viewing priority distribution as the outcome of a judgment on task utility, which was a recurring theme in the verbal reports of Experiments 1 and 3. In general, people are known to only engage in behavior if the rewards associated with that behavior (e.g., enjoyment) outweigh the predicted energetical costs (e.g., mental effort) (Kurzban et al., 2013; Boksem & Tops, 2008; Killu et al., 1999). Accordingly, the predicted energetical costs will have outweighed the limited rewards in Experiment 2. However, in the second dual-task condition of Experiment 3 the energetical costs have likely been lower, due to increased task-switching efficiency (Liepelt et al., 2011). Consequently, the evaluation of energetical costs and rewards has turned out favorably towards following the instructions in Experiment 3.

3.5.2 Integrated model for task prioritization

Until now, the switching mechanism of TCT has assumed fixed goal priority levels (see Figure 3.1). If, however, preferences cause variability in priority distribution, and if preferences are the result of utility judgments, then the next question is how to link such judgments with TCT. Task performance has been related with cost-benefit mechanisms (i.e., utility judgments) in several theoretical accounts (Hockey, 1997; Kurzban et al., 2013; Langner & Eickhoff, 2013; Sanders, 1983). The Compensatory Control Model (CCM) (Hockey, 1997; 2011), for example, describes the regulation of action in terms of a cost-benefit decision about the use of effort and the relative value of different goals. The higher one values a goal, the greater the willingness to spend additional effort on the corresponding task when its demands increase. An illustration of this cost-benefit decision is found in Experiment 2, where the driving task was protected against performance degradation, at the cost of decreased memory performance and increased mental effort.

We assume that cost-benefit decisions take place at a slower rate than the rapid switching mechanism described by TCT, analogous to the ‘slow’ and ‘fast’ systems of Kahneman (2011). Contrary to other cognitive-energetic models, the CCM allows for an explicit temporal distinction by capturing the above regulatory process in two control loops. Figure 3.12 describes a preliminary integration of TCT within CCM. The upper control loop features a cost-benefit decision structure, which adjusts goal priority levels in the goal buffer. In the lower control loop, TCT is modeled as a goal oscillator that switches between goals, as prescribed by Figure 3.1. In line with CCM, the goal oscillator adapts its output by comparing overt performance with the selected goals from the goal buffer. The lower control loop ‘sees’ goal priority levels in the goal buffer as constants, even though they are occasionally adjusted by the upper control loop. Thus, the control loops in this integration operate in different time domains.

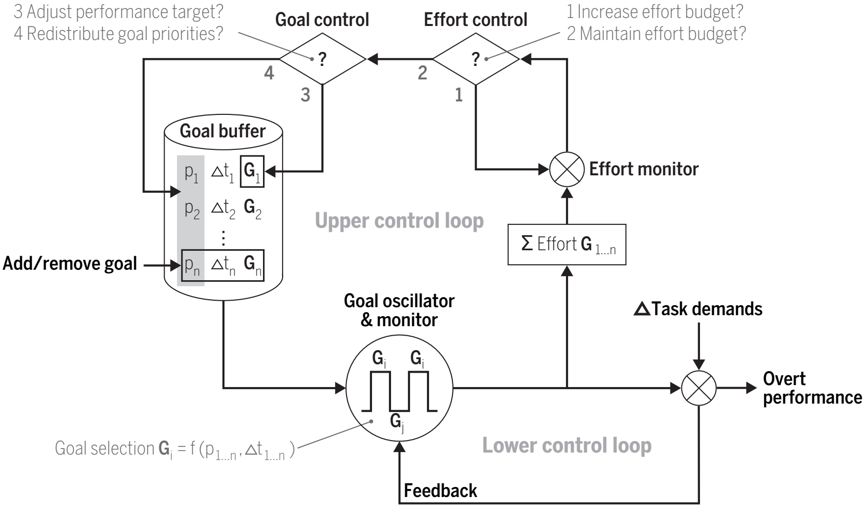

Figure 3.12. Integration of Threaded Cognition Theory (Salvucci & Taatgen, 2008) as goal oscillator within the Compensatory Control Model (adapted from Hockey (2011) with permission).

The adjustment of goal priority levels works as follows. An effort budget is used to compensate for sudden demand increments and resource decrements. The effort monitor compares the effort budget with the total effort level associated with the execution of all task goals. Therefore, the model in Figure 3.12 includes a summation of effort over iterations of partial task goal executions in the lower control loop. If the effort budget is insufficient to compensate for a discrepancy between intended performance and actual performance (e.g., failure to drive an intended route), CCM predicts a series of options (Hockey, 2011). The effort budget is either strategically raised to protect performance at the cost of fatigue (1), or lowered to prevent fatigue at the cost of task performance (2). These strategies are found with the 'equal' and 'driving' instructions in Experiment 3, respectively. Task performance decrements are either effectuated by adjusting the performance target of the current goal (3), or by displacing the current goal with a competing goal. We interpret goal displacement as a redistribution of priority levels (4). Ideally, priority instructions have a large effect on priority (re)distribution. Deviations from this ideal distribution are found when preferences result in an alternative cost/benefit decision.

3.5.3 Implications and future research

This explorative study provides several starting points for future research. From a theoretical perspective, a validation is needed of the proposed integration of TCT within CCM. We acknowledge that the proposed integration is currently not detailed enough to be implemented in the cognitive architecture in which TCT is modeled. However, recent studies show promising attempts at predicting single-task effort (Cao & Liu, 2011; Park & Myung, 2013), which provide an opportunity to test how effort drives task prioritization in concurrent multi-tasking. Specifically, these attempts may address the summation of effort in Figure 3.12, which features a transition from a fast process (e.g., goal oscillator, TCT) to a slow process (e.g., effort and goal control).

From a methodological perspective, the consequence of asking people afterwards about their preference, is that this procedure may result in unequal sample sizes, and low numbers in certain conditions. We acknowledge that this occurred in the present study. Indeed, when viewed per experiment, a low n may have reduced the reliability of the observed patterns. Looking across the experiments, however, we have observed several consistent patterns, yielding confidence in our overall results. For example, the distributions of the 'driving' and 'equal' preferences were consistent across Experiments 1 and 3. This helped to interpret the results of Experiment 2.

The question remains how to prevent unequal samples sizes when inquiring individual preferences. Asking people about their preferences beforehand is not a straightforward solution, because it may bias performance later on. Therefore, participant selection in future research benefits from having an inconspicuous method to predict preferences. If such method would exist, then knowledge on the likely distribution of preferences may prove instrumental in determining how many prospective participants should be recruited.

The causal role of preference on task prioritization was established through logical deduction from the combined results of Experiments 1 through 3. However, this deduction does not exclude other interpretations, such as the possiblity that participants assess their own performance, and then base their preference report on that. This issue, too, may be resolved by a method to predict preferences.

This study questioned the widespread assumption that people follow priority instructions in a dual-task setting. The assumption appears to be correct, provided that enough dual-task exposure is provided beforehand. A practical question, then, is exactly how much dual-task exposure is required before a conflicting priority instruction ‘wins’ against preference, and to what extent this is task- and context-dependent. In the traffic context, optimal safety requires drivers to prioritize the driving task at all times. This premise is not feasible for police officers, due to the dominant role of radio communication (Jansen et al., 2014; Sørensen & Pica, 2005). Although Dutch police officers do receive special driving training, they have to learn in the field how to balance between driving and listening. The present study suggests that these officers benefit from dual-task training to meet the implicit ‘equal’ priority instruction of police work, especially if this instruction conflicts with their task prioritization preferences.

Acknowledgments

We thank Hedderik van Rijn, Michael Steinborn, Niels Taatgen, and one anonymous reviewer for their many helpful comments and suggestions. This research was partially funded by Staf Korpsleiding / Directie Facility Management, Nationale Politie Nederland. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Altmann, E.M. & Trafton, J.G. (2002). Memory for goals: an activation-based model. *Cognitive Science 26*(1): 39-83.

Aretz, A.J. & Wickens, C.D. (1992). The Mental Rotation of Map Displays. *Human Performance 5*(4): 303-328.

Boksem, M.A.S. & Tops, M. (2008). Mental fatigue: costs and benefits. *Brain Research Reviews 59*(1): 125-139.

Bootes, K. & Chapparo, C. (2010). Difficulties with multitasking on return to work after TBI: a critical case study. *Work (Reading, Mass.) 36*(2): 207-216.

Caird, J.K., Willness, C.R., Steel, P. & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention 40*(4): 1282-1293.

Cao, S. & Liu, Y. (2011). Mental Workload Modeling in an Integrated Cognitive Architecture. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 55*(1): 2083-2087.

Cnossen, F., Meijman, T. & Rothengatter, T. (2004). Adaptive strategy changes as a function of task demands: a study of car drivers. *Ergonomics 47*(2): 218-236.

Dressel, J. & Atchley, P. (2008). Cellular phone use while driving: A methodological checklist for investigating dual-task costs. *Transportation Research Part F: Traffic Psychology and Behaviour 11*(5): 347-361.

Gopher, D. & Navon, D. (1980). How is performance limited: Testing the notion of central capacity. *Acta Psychologica 46*(3): 161-180.

Gopher, D., Weil, M. & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica 71*(1-3): 147-177.

Hancock, P.A. & Warm, J.S. (1989). A Dynamic Model of Stress and Sustained Attention. *Human Factors 31*(5): 519-537.

Hembrooke, H. & Gay, G. (2003). The laptop and the lecture: The effects of multitasking in learning environments. *Journal of Computing in Higher Education 15*(1): 46-64.

Hockey, G.R.J. (2011). A motivational control theory of cognitive fatigue. In: Ackerman, P.L. (Ed.) *Cognitive fatigue: multidisciplinary perspectives on current research and future applications*. Washington, DC, American Psychological Association: 167-188.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology 45*(1-3): 73-93.

Jansen, R.J., Van Egmond, R., De Ridder, H., & Silvester, S. (2014). Transitional Journey Maps: Capturing the dynamics of operational policing. In: De Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W., & Toffetti, A. (Eds.) *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*, 15-27, Torino, Italy.

Janssen, C.P., Brumby, D.P. & Garnett, R. (2012). Natural Break Points: The Influence of Priorities and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making 6*(1): 5-29.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, Farrar, Straus and Giroux.

Killu, K., Clare, C. & Im, A. (1999). Choice vs. Preference: The Effects of Choice and No Choice of Preferred and Non Preferred Spelling Tasks on the Academic Behavior of Students with Disabilities. *Journal of Behavioral Education 9*(3-4): 239-253.

Kurzban, R., Duckworth, A., Kable, J.W. & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences 36*(6): 661-679.

Langner, R. & Eickhoff, S.B. (2013). Sustaining Attention to Simple Tasks: A Meta-Analytic Review of the Neural Mechanisms of Vigilant Attention. *Psychological Bulletin 139*(4): 870-900.

Lansdown, T.C., Brook-Carter, N. & Kersloot, T. (2004). Distraction from multiple in-vehicle secondary tasks: vehicle performance and mental workload implications. *Ergonomics 47*(1): 91-104.

Levy, J. & Pashler, H. (2008). Task prioritisation in multitasking during driving: opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology 22*: 507-525.

Liepelt, R., Strobach, T., Frensch, P. & Schubert, T. (2011). Improved intertask coordination after extensive dual-task practice. *The Quarterly Journal of Experimental Psychology 64*(7): 1251-1272.

Miller, J. & Durst, M. (2014). 'Just do it when you get a chance': the effects of a background task on primary task performance. *Attention, Perception, & Psychophysics 76*(8): 2560-2574.

Nijboer, M., Taatgen, N.A., Brands, A., Borst, J.P. & van Rijn, H. (2013). Decision Making in Concurrent Multitasking: Do People Adapt to Task Interference? *PLoS ONE 8*(11): e79583.

Norman, D.A. & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology 7*(1): 44-64.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In: Boff, K.R., Kaufman, L. & Thomas, J.P. (Eds.) *Cognitive Processes and Performance*. John Wiley and Sons.

Park, S. & Myung, R. (2013). Predicting Task-related Properties of Mental Workload with ACT-R Cognitive Architecture. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 57*(1): 773-777.

Pashler, H. (1994). Dual-Task Interference in Simple Tasks: Data and Theory. *Psychological Bulletin 116*(2): 220-244.

Salvucci, D.D., Taatgen, N.A. & Borst, J.P. (2009). Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA, USA, ACM**:** 1819-1828.

Salvucci, D.D. & Beltowska, J. (2008). Effects of Memory Rehearsal on Driver Performance: Experiment and Theoretical Account. *Human Factors 50*(5): 834-844.

Salvucci, D.D. & Taatgen, N.A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review 115*(1): 101-130.

Sanders, A.F. (1983). Towards a model of stress and human performance. *Acta Psychologica 53*(1): 61-97.

Schultz, M. (2012). *RC Mini Racers*. Producer: Decane. Available: http://www.decane.net/game/mac- app-store/rc-mini-racers/. Accessed 1 December 2012.

Siu, K.-C. & Woollacott, M.H. (2007). Attentional demands of postural control: The ability to selectively allocate information-processing resources. *Gait & Posture 25*(1): 121-126.

Sørensen, C. & Pica, D. (2005). Tales from the police: Rhythms of interaction with mobile technologies. *Information and Organization 15*(2): 125-149.

Ünal, A.B., Platteel, S., Steg, L. & Epstude, K. (2013). Blocking-out auditory distracters while driving: A cognitive strategy to reduce task-demands on the road. *Accident Analysis & Prevention 50*: 934-942.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors 50*(3): 449-455.

Wickens, C.D. (1984). Processing resources in attention. In: Parasuraman, R. & Davies, D.R. (Eds.) *Varieties of attention*. London, Academic Press: 63-102.

Zar, J.H. (1996). *Biostatistical Analysis*, 3rd ed. Upper Saddle River, New Jersey: Prentice Hall.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior - A design approach for modern tools*. Delft, The Netherlands, Delft University of Technology. PhD thesis.

Chapter 4

Impact of task prioritization on coping strategies under varying task demands

Abstract

This study investigates how task prioritization influences coping behavior under varying task demands. As a first step, we postulate the logics to infer the coping strategies of Hockey's (1997) Compensatory Control Model from tradeoff patterns in dual-task performance and effort, with explicit attention to task prioritization. Two new coping strategies follow from these logics, labelled as 'intense focus' and 'exclusive decrement'. The second step concerns a dual-task experiment with an auditory memory task and a driving task, based on the context of police work. Task demands were manipulated through signal-to-noise ratio and route curvature. For each of two priority instructions (driving, equal), coping strategies were inferred through pair-wise comparisons between the experimental conditions. Expected coping strategies were found in comparisons between single-task and dual-task conditions, but not in comparisons between pairs of dual-task conditions. Furthermore, none of the comparisons yielded an

identical coping strategy for both instructions. Therefore, it is essential to involve task prioritization in dual-task driving studies.

4.1 Introduction

> *The dispatcher calls: 'Attention unit 5030, we received report of a molestation.' Police officers L. and A. register the incoming message through their portophone ear pieces. L. reflects on the auditory demands of radio communication in police work: 'On my left side I am deaf to other sounds because of that ear piece. When I am at home after a busy shift I still hear the voices.' A. responds: 'Sometimes the portophone distorts, and you cannot properly hear what is being said.' When A. starts driving at 160 km/h, the light bar on top of the vehicle emits a strong noise due to the increased air flow. 'Can you still hear the portophone over the speakers? We can barely understand each other.' L. nods: 'And imagine what it is like when you activate the siren.'* (field notes from Jansen et al., 2014)

The above example illustrates the high demand level of radio communication, which plays a dominant role in police work (Anderson et al., 2005; Sørensen & Pica, 2005). The demands of such in-vehicle tasks are known to compromise safe driving behavior (Caird et al., 2008; Dingus et al., 2016; Lee et al., 2001; Strayer et al., 2003). An additional safety hazard of police work is found in the excessive driving speeds during emergency responses and pursuits. High driving speed has not only been related to higher accident rates with regular (i.e., non-police) drivers (Aarts & Van Schagen, 2006), but it has also been shown to be an explanatory factor for accidents involving police officers (Clarke et al., 2009). These safety hazards suggest that police officers must employ different strategies than regular drivers to cope with situational task demands. This study investigates such coping strategies in terms of tradeoffs between task performance and mental effort (cf. Hockey, 1997, 2011).

Contrary to regular drivers, police officers on solo patrol typically do not have the option to stop their car to attend important incoming messages, nor can they afford an uninformed arrival at the scene. In other words, regular drivers are expected to fully prioritize the driving task, whereas police officers are expected to prioritize the driving and radio communication tasks with approximately equal priority. The distinction between primary and secondary tasks thus becomes ambiguous when drivers adopt additional social roles (cf. Hancock et al., 2008). In light of the safety hazards above, the present study investigates how task prioritization (i.e., the process of allocating attention to one task at the expense of another task) influences the selection of coping strategies under varying task demands.

When drivers are confronted with increased task demands (e.g., a sudden phone call, increased road curvature), they have been known to protect the highest priority task goal (e.g., arriving safely) by investing more effort, or by adjusting the performance targets associated with each task goal (e.g., accepting a later time of arrival by lowering driving speed). Hockey's (1997, 2011) Compensatory Control Model (CCM) accounts for this ability by predicting a number of coping strategies. Surprisingly, previous studies that were comparable in experimental setups have resulted in the inference of different coping strategies. In this paper, we present a set of methodological requirements to infer those coping strategies from tradeoffs between task performance and effort. The requirements have been implemented in a dual-task experiment, in which differences between police officers and regular drivers have been emulated through priority instructions.

4.1.1 Factors influencing the selection of coping strategies

Hockey (1997) observed that people follow different strategies to maintain primary task performance under conditions of increasing task demand, at the expense of secondary task performance and/or effort. His Compensatory Control Model (CCM) (Hockey, 1997, 2011) summarizes five coping strategies. First, the 'compensatory costs' strategy refers to investing more effort to meet the performance targets of all tasks. Second, 'secondary decrement' corresponds with stabilizing primary task performance and effort at the cost of secondary task performance (e.g., paying less attention to radio items). Third, ‘strategic adjustment’ is about stabilizing effort and secondary task performance by shifting to simpler strategies within the primary task (e.g., accepting a later time of arrival by lowering driving speed). Fourth, ‘fatigue after-effects’, is a preference for tasks with low work demands after prolonged exposure to high work demands. This strategy is typically preceded by the ‘compensatory costs’ strategy. Fifth, ‘disengagement’ from the pursuit of all task goals may occur to protect energetic resources. With the exception of ‘fatigue after-effects’, these coping strategies manifest themselves as direct tradeoffs between task performance and effort.

Given that multiple coping strategies exist, then what initiates the selection of these strategies? Cnossen et al. (2004) suggests that in dual-task contexts the source of increased task demands (i.e., primary vs. secondary tasks) influences the selection of coping strategies. This suggestion was based on an experiment with regular drivers, in which the driving task was assumed to be protected. When traffic density increased (i.e., a manipulation of primary task demand), driving speed decreased, whereas no effects were found on memory performance, nor on mental effort. This tradeoff was interpreted in terms of ‘strategic adjustment’. In contrast, an increase in the demands of a secondary navigation task resulted in lower memory performance and increased mental effort. Cnossen et al. (2004) interpreted this tradeoff in terms of the ‘secondary decrement‘ coping strategy.

Support for the suggestion of Cnossen et al. (2004) is inconclusive. On the one hand, Horberry et al. (2006) found distinct tradeoff patterns between performance and mental workload by manipulating the demands of the primary task (i.e., driving) and the secondary task (i.e., phone conversation). Increasing traffic density and visual complexity of the driving environment resulted in lower driving speed, but did not affect secondary task performance, nor mental workload. On the other hand, the addition of a phone conversation did not affect driving performance, but did result in increased mental workload. Thus, the manipulations of primary and secondary task demands can be interpreted in terms of ‘strategic adjustment’ and ‘compensatory costs, respectively.

On the other hand, the results of a study by Ünal et al. (2013) challenge the validity of Cnossen et al.’s suggestion. The addition of a memory task to a driving task did not affect driving performance, but compared to single-task memory performance, less items were recalled. Increasing traffic density yielded similar results: no effect on driving performance, and decreased memory performance. Thus, manipulations of primary and secondary task demands yielded the same coping strategy (i.e., ‘secondary decrement’). In conclusion, Ünal et al.’s (2013) inconsistency with Cnossen et al. (2004) and Horberry et al. (2006) suggests that the source of increased task demand is not the decisive factor that influences coping behavior.

We hypothesize that coping behavior depends not only on the source of increased task demand, but also on task prioritization. This hypothesis is guided by the expected difference in task prioritization between police officers and regular drivers. Task prioritization is known to affect tradeoffs in task performance (Gopher & Navon, 1980; Norman & Bobrow, 1975), as well as tradeoffs between performance and mental workload (Gopher & Donchin, 1986; Tsang et al., 1996). In this light it should be noted that neither Cnossen et al. (2004), nor Horberry et al. (2006), nor Ünal et al. (2013) have systematically manipulated task prioritization. In fact, Cnossen et al. (2004) point out that performance decrements may be adequate from the driver’s perspective if the priorities of task goals are set accordingly. Thus, if participants differed in their judgment on the relative importance of each of those task goals (cf. Jansen et al., 2016), this could explain why different tradeoff patterns have been found, and consequently, different coping strategies. An obvious solution to this problem is to provide explicit priority instructions (Dressel & Atchley, 2008; Janssen et al., 2012), which, as will be discussed next, is one of our proposed methodological requirements to infer coping strategies.

4.1.2 Requirements for inference of coping strategies

The above inconsistency in inferred coping strategies warrants a closer look into the relation between coping strategies, task prioritization, and tradeoff patterns. Hockey's (1997, 2011) CCM describes the management of effort to regulate task performance depending on the relative importance of competing task goals, changes in task demands, and current levels of energetic resources. Two control mechanisms are available in case of a discrepancy between overt task performance and an internally represented task goal. First, effort, which Hockey (2011) views as an optional response to the perception and appraisal of demands, may be increased to protect the task goal. As a second control mechanism, performance targets may be reduced to maintain the current effort level, thereby preventing fatigue buildup in the long run. Finally, Hockey (1997) states that task prioritization steers the tradeoff patterns associated with coping strategies.

From the perspective of the driver, the relation between coping strategies, task prioritization, and tradeoff patterns can be captured as follows for our dual-task context:

$$\begin{gathered}\text{Coping strategy}_{\text{i}} \mid \text{Task prioritization}_{\text{j}} \rightarrow \\ \{\Delta\text{Performance}_{\text{Driving}}, \Delta\text{Performance}_{\text{Communication}}, \Delta\text{Effort}\}\end{gathered} \tag{4.1}$$

where *i* represents any of the above coping strategies, excepting 'fatigue after-effects', and *j* represents the proportional distribution of attention between the tasks. Changes in performance and effort, then, are a direct consequence of the control mechanisms. How, then, can coping strategies be inferred from behavior? Formula 4.2 represents the above relation from the perspective of the experimenter:

$$\begin{gathered}\{\Delta\text{Performance}_{\text{Driving}}, \Delta\text{Performance}_{\text{Communication}}, \Delta\text{Effort}\} \mid \\ \text{Task prioritization}_{\text{j}} \rightarrow \text{Coping strategy}_{\text{i}}\end{gathered} \tag{4.2}$$

Formula 4.2 suggests that the inference of a coping strategy concerns an interpretation of a performance/effort tradeoff with a given task prioritization. The CCM describes the effects of coping strategies on task performance in terms of either decrements, or no decrements. Effects on mental effort are described in terms of either increments, or no increments. As a result, dual-task performance/effort tradeoffs can be organized into eight logical combinations, as presented in the first three columns of Table 4.1.

The fourth column of Table 4.1 concerns the case of regular drivers, where a primary task (i.e., driving) is prioritized over a secondary task (i.e., phone communication). Four tradeoffs correspond with CCM coping strategies, whereas the other tradeoffs require an alternative interpretation. Hockey &

Table 4.1. Coping strategies inferred from changes in task performance and mental effort in a dual-task context, and as function of task prioritization.

ΔPerf driving	**ΔPerf communication**	**ΔEffort**	**Coping strategy (‘driving’ priority)**		**Coping strategy (‘equal’ priority)**	
ND	ND	NI	N.A.		N.A.	
ND	ND	I	YES	Compensatory costs	YES	Compensatory costs
ND	D	NI	YES	Secondary decrement	YES	**Exclusive decrement**
ND	D	I	YES	**Intense focus**	NO	Failed regulatory control
D	ND	NI	YES	Strategic adjustment	YES	**Exclusive decrement**
D	ND	I	NO	Failed regulatory control	NO	Failed regulatory control
D	D	NI	YES	Disengagement	YES	Disengagement
D	D	I	NO	Failed regulatory control	NO	Failed regulatory control

NOTE: ND = no decrement, D = decrement, NI = no increment, I = increment. New coping strategies are represented in bold (based on: Hockey, 1997, 2011; Hockey & Earle, 2006).

Earle (2006) interpret decreased performance with increased effort as ‘failed regulatory control’. However, this interpretation is not justifiable for the combination of stable primary task performance, decreased secondary task performance, and increased effort, because the primary task is protected against degradation. One could argue that this tradeoff fits with the strategy of ‘secondary decrements’, but this interpretation does not comply with the fact that there are also ‘compensatory costs’. We label the combination of protected primary task performance, secondary decrements, and compensatory costs as ‘intense focus’. This corresponds with the situation in which primary task demand is so high, that despite neglecting the secondary task, additional effort is still required.

Although the CCM assumes one highest priority task goal, recent work by Jansen et al. (2016) suggests that the CCM can be adapted to account for situations with multiple high priority goals. The fifth column of Table 4.1 represents the case of police officers, where two tasks have equal priority. This implies that one can no longer speak of ‘primary’ and ‘secondary’ tasks, and the labels ‘strategic adjustment’ (e.g., referring to a primary task) ‘secondary decrement’ (e.g., referring to a secondary task) are no longer applicable. Given that both performance/cost tradeoffs concern a performance decrement in only one task, we label both cases as ‘exclusive decrement’. In addition, the performance/cost tradeoff previously labelled as ‘intense focus’ should now be interpreted as ‘failed regulatory control’, because the redistribution of priorities no longer justifies such a focus on one task.

Based on Formula 4.2 and Table 4.1, three requirements should be met for a proper inference of coping strategies in an experimental dual-task setting. First, a manipulation of task demand is required to induce coping behavior (e.g., from single-task to dual-task, or through difficulty levels within a dual-task setting). Second, performance and effort data should be collected for both tasks on both demand levels for an overview of potential tradeoffs (Dressel & Atchley, 2008; O’Donnell & Eggemeier, 1986). Third, unambiguous task prioritization is needed to interpret these tradeoffs, as demonstrated by the alternative interpretations of three performance/

effort tradeoffs in Table 1. Surprisingly, it is not common practice to report secondary task baseline data when describing coping behavior at single-to-dual task transitions (e.g., Rakauskas et al., 2004; Törnros & Bolling, 2005). This is likely because of a focus on driving safety from the perspective of regular drivers, as opposed to the multi-tasking context of police officers. Another re-occurring methodological shortcoming in the interpretation of tradeoffs is the absence of explicit priority instructions (e.g., Haigney et al., 2000; Cnossen et al., 2004; Rakauskas et al., 2004; Ünal et al., 2013). These issues are addressed in the present study.

4.1.3 Paradigm

The present study examines the influence of task prioritization and source of increased task demand on coping behavior. Coping behavior was not measured directly, but inferred from measurable task behavior and effort ratings. Inspired by the context of solo patrol in operational policing (Jansen et al., 2014), a dual-task setting has been used with a driving task and an auditory memory task. The driving task required a high driving speed level to simulate police emergency response driving. Driving task demand has been manipulated through road curvature, in that drivers had alternated between following a straight route and a curvy route (cf. Alm & Nilsson, 1994; Jamson & Merat, 2005). Both routes were trained to minimize the navigation component of the driving task, thereby minimizing potential order effects in subsequent experimental conditions. The memory task was continuous, corresponding with the on-going demands of attending police radio messages. The stimuli concerned news fragments, based on previous work by Jansen et al. (2016). The anecdote in the introduction illustrates how distortion in the portophone signal and interference by wind noise both place heavy demands on the police officers. Baldwin (2007) argues that lowering the signal-to-noise ratio of a message will increase the difficulty to comprehend its content. Therefore, two stimulus versions have been created to manipulate the demands of the memory task: distorted, and clean (i.e., non-distorted).

Task prioritization has been manipulated through separate priority instructions to address the difference between regular drivers (i.e., a 'driving' instruction) and police officers (i.e., an 'equal' instruction). There is a potential risk in using priority instructions, though, that should be accounted for. Our previous work (Jansen et al., 2016) suggests that people are not always able or willing to follow instructions, because they have distinct preferences regarding task prioritization (i.c., 'driving' and 'equal'). It has been shown that the effect of such conflict decreases through additional dual-task exposure, which is believed to improve the efficiency of rapid task switching (Liepelt at al., 2011). In the present study, we have taken three measures to minimize the effect of potentially conflicting instructions. First, the instructions are based on naturally occurring task prioritization preferences in our previous work (Jansen et al., 2016), to ensure that the instructions were non-conflicting at least half of the time. Second, task prioritization has been implemented as within-subjects factor to increase dual-task exposure. Third, by minimizing the navigation component of the driving task, less frequent task switching is required between the driving task and the memory task. This should reduce the amount of dual-task exposure that is required to overcome a conflicting instruction.

From prior work we expect that each priority instruction will yield distinct tradeoff patterns, and therefore, distinct coping strategies. For a police officer, the only way to safe-guard an informed and fast arrival at the scene seems to invest more effort on both tasks. However, prolonged investment of effort leads to fatigue in the long run (Hockey, 2011). This side effect was demonstrated by officers' complaints about fatigue as a result of coping with the above task combination (Jansen et al., 2014). Therefore, the 'equal' instruction in the present study is expected to induce the 'compensatory costs' coping strategy, regardless of the source of increased task demand. Regarding the 'driving instruction, we expect an occurrence of the 'strategic adjustment' and/or 'secondary decrement' coping strategies, in line with previous studies on regular drivers (Cnossen et al, 2004; Horberry et al., 2006; Ünal et al., 2013). However, no predictions can be made regarding the source of increased task demand, due to the inconsistencies described above.

We have investigated two types of transitions between the experimental conditions. First, transitions between dual-task conditions have been analyzed to test the hypothesis that coping behavior is influenced both by task prioritization and by the source of increased task demand. Second, transitions between single-task and dual-task conditions have been analyzed to exclusively investigate the influence of task prioritization. For the latter transition type it is not possible to test the hypothesis, due to an absence of a clear source of increased task demands. One may argue that in case of the ‘driving’ instruction, the secondary memory task is added to the primary driving task, thus regarding the memory task as source of increased task demand. However, such a rationale does not hold for the ‘equal’ instruction, because the distinction between ‘primary’ and ‘secondary’ tasks does not apply.

Our approach was to first verify that the priority instructions were followed, and subsequent analyses were performed for each instruction separately. The reason for this approach is that the CCM predicts coping strategies based on changes in task demands, but not based on changes in task prioritization (e.g., from ‘driving’ to ‘equal’). The state transition diagram in Figure 4.1 represents the dual-to-dual comparisons that can be made within each instruction, in which the direction of each arrow corresponds with the expected increase in task demands. Note that no directional arrow is drawn between ‘curvy/clean’ and ‘straight/distorted’, because it is unknown which of the two manipulated factors (i.e., Route type, Signal type) has the largest effect on overall task demands. The arrows in Figure 4.1 have been used to make planned pair-wise comparisons. Finally, Table 4.1 has been used to interpret significant tradeoffs between performance and mental effort in terms of coping strategies.

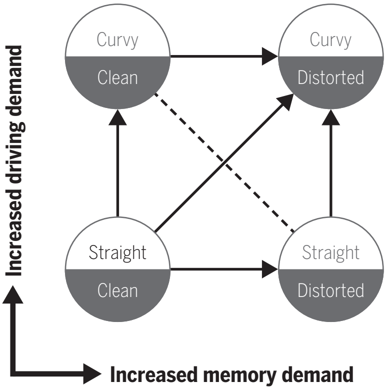

Figure 4.1. State transition diagram with comparisons between pairs of dual-task conditions. Solid arrows represent an increase in task demand. The dashed line indicates uncertainty about the direction of increased demand.

4.2 Method

4.2.1 Participants

Twenty-four students of the Faculty of Industrial Design Engineering and the Faculty of Mechanical Engineering at Delft University of Technology (DUT) participated in the experiment (24 males, 18 to 27 years old, average 22.2 years). This study was approved by the Ethical Committee of DUT. All participants reported experience with playing action games (e.g., racing, first person shooter). They were native Dutch speakers, reported normal hearing, and normal or corrected-to-normal vision. Participants received a compensation of 10 euros at the end of the experiment.

4.2.2 Auditory memory task

The goal of the memory task was to correctly answer a question for each auditory stimulus. Forty-eight stimuli were prepared, of which eight were used for training. Stimuli consisted of a Dutch news items (average duration: 15 sec), spoken by professional male news readers. News items were at least half a year old to minimize recognition. A factual question was recorded for each news item by a native speaker from The Netherlands. Questions were related to information around the center of a news item to minimize primacy and recency effects (Acheson & MacDonald, 2009), and allowed for one correct answer only. For example, the item: *"In the third quarter of this year less cars were sold than in the same period of last year. To be precise: six percent less. The trade organizations also expect a decrease in sales next year." was accompanied by the question: "How many percent less cars were sold?"*.

A second set of forty-eight auditory stimuli was created by applying a distortion effect to the above news items. To mimic a police radio context, this effect involved compression, the addition of hailstorm noise recorded inside a car, signal clipping, and band-pass filtering (300-3400Hz, -36dB/oct). The amount of distortion was varied for each news item. The signal-to-noise ratio was measured through harmonicity in Praat (Boersma & Weenink, 2013) for the maximum amount of distortion at which news items were still comprehensible in a pilot test. Harmonicity was significantly lower for distorted news items (M = 2.60, SE = .073), compared to the original versions (M = 7.29, SE = .17), $t(47)$ = 38.95, $p < .001$, r = .75. In a second pilot test, ratings on the Rating Scale Mental Effort (Zijlstra, 1993) confirmed that answering questions on distorted stimuli was more demanding (M = 91.86, SE = 10.66) than without distortion (M = 56.17, SE = 9.05), $t(11)$ = -2.50, $p < .029$, r = .60. Volumes across all news items and questions were matched. Stimuli and questions were saved as wav files (16 bit, 44.1 kHz).

4.2.3 Driving task

The 'RC Mini Racers' (Schultz, 2012) game was used, in which a miniature vehicle was controlled by the arrow keys. The vehicle was situated in a closed virtual environment without moving objects. Two routes were designed within this environment: 'straight' and 'curvy'. The straight route imposed a relatively low task demand, because it consisted of unobstructed straight road sections with gradual turns. The curvy route, on the other hand, imposed a relatively high task demand, because the straight road sections were shorter, and because its sharp turns required the driver to avoid suddenly appearing objects. The lengths of the routes were chosen such that their completion took a similar amount of time, which was confirmed with an informal pilot. Given that the sharp turns of the curvy route necessitated a greater speed reduction than the gradual turns of the straight route, the straight route was made 1.3 times as long as the curvy route. The goal of the driving task was to repeatedly drive either route as many times as possible within the duration of an experimental condition (200 sec). A button had to be pressed to return to the starting location each time the destination was reached.

4.2.4 Experimental design

Four factors were varied in the experiment: the curvature of the driving task (Route: straight versus curvy), the signal-to-noise ratio in the memory task (Signal: clean versus distorted), task prioritization (Instruction: ‘driving’ versus ‘equal’), and task composition (Task: single versus dual). A mixed design was used, with Signal as between-subjects factor, and Route, Instruction, and Task as within-subjects factors. As a result, an experimental session consisted of three single-task conditions (i.e., memory baseline, straight route baseline, curvy route baseline), and four dual-task conditions (i.e., ‘driving/straight’, ‘driving/curvy’, ‘equal/straight’, and ‘equal/curvy’).

4.2.5 Measures

Analogous to emergency response time in operational policing, driving performance was measured in terms of route duration (e.g., Jordan & Johnson, 1993). Only attempts in which a route was fully finished were considered. Driving performance within each condition was calculated as the average duration of all successful attempts. Memory performance within each condition was calculated as the proportion of correct answers. An arcsine transformation on the proportional scores (Zar, 1996) was used for subsequent analyses. Subjective mental effort was obtained through a Dutch version of Zijlstra’s (1993) Rating Scale Mental Effort (RSME).

4.2.6 Apparatus

The experiment took place in a well-lit, quiet room. The driving task was played on an Apple MacBook Pro 15”. Each route was printed on an A3 sheet of paper, and placed next to the laptop. A separate button box was constructed to facilitate a swift response at the end of a route. A pair of Creative Gigaworks T20 Series II loudspeakers were used for playback of the auditory stimuli and the driving game sounds. The loudspeakers were positioned at ear height, approximately 30 cm to the left and right of the laptop. A dedicated program, coded in Max v.6 (Cycling74, Inc.), was used for randomization of the stimuli across experimental conditions, for playback of the stimuli, to record verbal responses, to measure route durations, and to collect RSME ratings.

4.2.7 Procedure

After signing an informed consent form, participants were randomly assigned to ‘clean’ messages (n = 12), or to ‘distorted’ messages (n = 12). Sessions were organized in four phases: training, three single-task baseline conditions, four dual-task conditions, and an interview. Participants first habituated to the memory task with a separate set of 8 training trials (average duration: 25 sec, total duration: 200 sec.). Each memory trial was composed of a news item, a question, 4.5 sec answer time, and a beep sound. The participant was informed about this sequence, and instructed to verbalize an answer during the designated interval. An answer after the beep was allowed, but the participant was urged to prepare for the next trial. The volume was set to a comfortable listening level. Single-task baseline performance on the memory task was established by repeating this protocol with 8 stimuli, randomly selected from a pool of 40 test stimuli.

The participant was introduced to the controls of the car and the driving environment. Game sounds were included for feedback on driving speed, but their volume was set to a lower volume to ensure audibility of the stimuli in the subsequent dual-task conditions. The participant first received the map of the straight route. Single-task baseline performance on the straight route was established by completing this route as fast as possible at least five times, until a) the maximum duration of the last three attempts was within a four second margin of the minimum duration by that participant so far, and b) the last three attempt durations were not in descending order. Baseline performance was calculated as the average duration of the last three attempts. This protocol was repeated for the curvy route.

The remaining 32 test stimuli were randomly distributed over four dual-task conditions. The order of these conditions was counterbalanced through a Latin square design. The participant received a priority instruction before each condition. The 'driving' instruction was to prioritize the driving task over the memory task. Attending the memory task was allowed, as long as this would not degrade driving task performance (i.e., driving the route as fast and often as possible). The 'equal' instruction, on the other hand, required the participant to perform both tasks as well as possible. Mental effort was administered through the RSME after single-task baseline conditions and dual-task conditions. Sessions ended with a semi-structured interview.

4.2.8 Data analysis

All statistical tests were performed with SPSS v.20, and results were compared to an α level of .05. Two types of analyses were conducted. First, mixed 2 (Route) x 2 (Signal) x 2 (Instruction) ANOVAs were conducted on the dual-task conditions to evaluate whether the priority instructions were followed. Second, tradeoffs patterns between performance and mental effort were identified for each instruction separately. Planned pair-wise comparisons were made between dual-task conditions (six pairs in total, see Figure 4.1), as well as between single- and dual-task conditions (four in total). One-tailed *t*-tests were used for expected increases in task demand (i.e., single- to dual-task, and see arrows in Figure 4.1), and two-tailed *t*-tests when the direction of increased task demand was unknown (see dashed line in Figure 4.1). Finally, Table 4.1 was used to infer coping strategies from the identified tradeoffs.

4.3 Results

First, a manipulation check was performed to verify that the participant groups were equivalent, and that the within-subjects factors did not introduce order effects. Second, it was verified whether the priority instructions were followed. Third, tradeoff patterns between performance and mental effort were examined for each instruction separately. Finally, these tradeoff patterns were interpreted as coping strategies.

4.3.1 Manipulation check

Single-task baselines

Both signal type groups (i.e., clean, distorted) have performed the driving baseline task. To identify potential differences between these groups, we performed a 2 (Signal) x 2 (Route) mixed ANOVA on the driving baseline task. No significant effects were found on driving performance, nor on mental effort. This finding suggests that, by the end of the training protocol, the participant groups were equivalent regarding driving task experience. The interaction between Signal and Route was non-significant. However, significant main effects of Route were found on mental effort and driving performance. The curvy route (M = 55.30, SE = 4.46) required significantly more effort than the straight route (M = 40.75, SE = 4.27), $F(1,22) = 25.13$, $p < .001$, $\eta_p^2 = .53$, which reflects the intended manipulation of driving task demand. The straight route (M = 43.94s, SE = .34) took longer to complete than the curvy route (M = 37.78s, SE = .57), $F(1,22) = 142.31$, $p < .001$, $\eta_p^2 = .87$. This difference can be attributed to the length of the straight route being 1.3 times of the curvy one, and not to differences in mental effort.

In the memory baseline condition, no significant performance difference was found between clean stimuli (M = 58.33%, SE = 4.95) and distorted stimuli (*M* = 57.29%, *SE* = 7.61). This finding suggests both groups were equally skilled at the memory task. Furthermore, distorted stimuli required significantly more mental effort (*M* = 79.56, *SE* = 4.86) than clean stimuli (*M* = 67.78, *SE* = 3.84), *t*(22) = -1.90, *p* = .035, *r* = .38. This means the intended manipulation of auditory task demand was effective.

Averaged over signal type and route type, the separate memory task (*M* = 73.67, *SE* = 3.27) required significantly more effort than the separate driving task (*M* = 48.02, *SE* = 4.13), *t*(23) = 5.89, *p* < .001, *r* = .33. This finding suggests that in the dual-task conditions the memory task had a greater influence on effort than the driving task.

Dual-task order effects

Within the fixed duration of the dual-task conditions, participants typically performed four successful attempts on the straight route, and four to five successful attempts on the curvy route. Twenty participants reported that they did not look at the map during the dual-task conditions. One other participant only looked at the map during each first attempt after switching routes. This finding suggests that navigation had become an automated process, as was intended to minimize order effects. A 4x2 ANOVA was conducted on each dependent variable, with the order of occurring conditions as repeated factor. No significant order effects were found.

4.3.2 Overview experimental data

Figure 4.2 shows the main results per dual-task condition, organized as function of instruction, route type, and signal type (i.e., open vs. filled circles). In addition, single-task baseline data have been included, represented by dashed lines. The dual-task data were subjected to a 2x2x2 mixed ANOVA to verify if and how the priority instructions were followed, with Instruction (driving, equal) and Route (straight, curvy) as within-subjects factors, and Signal (clean, distorted) as between-subjects factor. The results of these tests (see Table 4.2) are discussed separately for each measure.

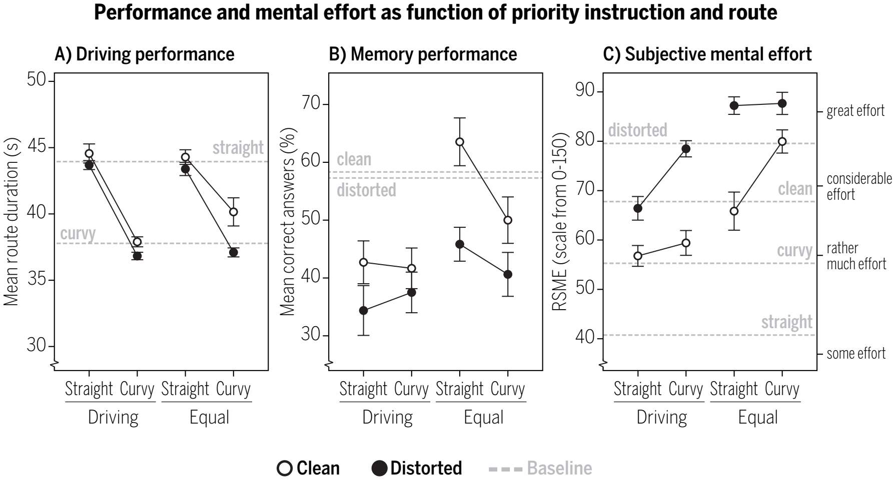

Figure 4.2. Driving performance (panel A), Memory performance (panel B), and subjective mental effort (panel C) as function of Route type, Signal type, and Instruction. Solid lines are added for interpretation only. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability.

Table 4.2. Summary of ANOVA results on performance and mental effort.

	Driving performance			**Memory performance**			**Mental effort**		
Source	*F*	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Instruction	1.32	.26	.057	**11.34**	**.003**	**.34**	**52.71**	**< .001**	**.71**
Route	**119.23**	**< .001**	**.84**	3.66	.070	.14	**14.16**	**.001**	**.39**
Signal	2.87	.11	.12	**4.80**	**.039**	**.18**	**4.52**	**.045**	**.17**
Instr x Route	3.36	.081	.13	2.04	.17	.085	< .001	.99	< .001
Instr x Sig	1.38	.25	.059	1.12	.30	.048	.002	.96	< .001
Route x Sig	1.16	.29	.050	1.62	.22	.070	.30	.59	.013
Instr x Route x Sig	1.34	.26	.058	.032	.86	.001	**8.41**	**.008**	**.28**

NOTE: *df* = (1,22). Instr = Instruction, Sig = Signal. Significant results in bold.

The mean route durations per dual-task condition are plotted Figure 4.2A. The instructions clearly did not affect driving performance. The graph does show that, consistent with single-task baseline performance, participants take significantly longer for the straight route than the curvy route.

The mean proportions of correct answers per dual-task condition are plotted in Figure 4.2B. Shifting priority from ‘driving’ to ‘equal’ had a positive and significant effect on memory performance. In addition, performance with distorted news items was significantly lower than with clean news items, which suggests that distortion inhibited memory performance. Finally, dual-task and single-task memory performance appear to be similar with the combination of the ‘equal’ instruction and a clean signal type.

The anchor words on the right side of Figure 4.2C show that the effort ratings mostly varied between ‘rather much effort’, and ‘great effort’. The ‘equal’ instruction required significantly more mental effort than the ‘driving’ instruction. The other main effects were also significant. Distorted news items required more mental effort than clean news items, and the curvy route required more mental effort than the straight route. In addition to these main effects, a significant three-way Priority x Route x Signal interaction was found. With the ‘driving’ instruction, mental effort appears relatively stable between the two routes when the news items are clean. With distorted news items, however, mental effort increases when the route becomes more difficult. The directions of these trends are reversed with the ’equal’ instruction: clean news items are accompanied by an increase in mental effort when route difficulty increases. However, mental effort with distorted stimuli is rated at what appears to be a maximum value, regardless of the route.

Before analyzing the tradeoffs between dual-task performance and effort separately for each instruction, it is necessary to establish that the priority instructions were actually followed. Figure 4.2B shows that memory performance is higher with the ‘equal’ instruction, but according to Figure 4.2C this comes at the cost of increased mental effort. This tradeoff between memory performance and mental effort is in line with our previous work on task prioritization (Jansen et al., 2016), and strongly suggests that the instructions were indeed followed. No significant main effect of Instruction was found on driving performance. This was to be expected, because protection of the driving task was the intention of both instructions.

4.3.3 Planned pair-wise comparisons

Tradeoff patterns were explored in detail through planned pair-wise comparisons between the dual-task conditions. This was done separately for each priority instruction, because they proved to result in distinct patterns of performance and effort. But first, a potential issue with the interpretation of driving performance needs to be addressed. Route durations in the single-task baseline conditions were significantly different, an effect which persisted from single- to dual-task conditions. If this difference was caused by the greater distance of the straight route, then the longer duration to complete the straight route, compared to the curvy route, cannot unambiguously be interpreted as a performance decrement. Consequently, a direct comparison between the average durations of the straight and curvy routes does not provide a clear insight into the effect of increased task demand on driving performance. We have used the following workaround. For each participant, dual-task driving performance was divided by the corresponding single-task baseline performance. The resulting proportional scores were subsequently used to compare driving performance across the dual-task conditions (see Tables 4.3 and 4.4). Table 4.4 shows that driving performance significantly improved (i.e., shorter relative duration) at the transition from 'curvy/clean' to 'curvy/distorted' with the 'equal' instruction. This is contrary to our prediction, because the latter condition imposes a higher demand level. As the effect was non-significant with a two-tailed test, it was considered as non-significant in subsequent analyses on tradeoff patterns. Using Cohen's (1988) guidelines on effect sizes, most significant effects in Tables 4.3 and 4.4 had medium (i.e., $r > .30$) to large (i.e., $r > .50$) effect sizes.

Tradeoffs between dual-task and the corresponding single-task baseline conditions were also examined. For example, performance and mental effort in the 'curvy/distorted' condition were compared with performance and mental effort in the 'curvy' and 'distorted' single-task conditions. Given that both single-task conditions obtained mental effort, tradeoffs can be established using either the driving task as baseline for mental effort, or

Table 4.3. t-test results for transitions between dual-task conditions for the 'driving' instruction. Dual-task driving performance on the straight route was divided by baseline performance on the straight route, and dual-task driving performance on the curvy route was divided by baseline performance on the curvy route. Str = Straight, Cur = Curvy, Cl = Clean, Dis = Distorted. One-tailed p values were used, except for the transition between 'curvy/clean' and 'straight/distorted'. Significant results in bold, * $p < .05$, ** $p < .01$.

Transition			**Δ Driving performance**			**Δ Memory performance**			**Δ Mental effort**		
From	**To**	**df**	***M***	**t**	**r**	***M***	**t**	**r**	***M***	**t**	**r**
Str_Cl	Str_Dis	22	-.011	.61	.13	-8.33	1.27	.26	9.64	1.26	.26
Cur_Cl	Cur_Dis	22	.0079	-.41	.087	-4.17	0.63	.13	19.06	2.43*	.46
Str_Cl	Cur_Cl	11	-.023	.40	.069	-1.04	.28	.36	2.63	-.73	.77
Str_Dis	Cur_Dis	11	-.0044	.35	.89	3.13	-.42	.24	12.06	-4.09**	.89
Str_Cl	Cur_Dis	22	-.015	.84	.18	-5.21	.86	.18	21.69	-3.10**	.55
Cur_Cl	Str_Dis	22	.012	-.65	.14	-7.29	1.02	.21	7.01	-.83	.17

Table 4.4. t-test results for transitions between dual-task conditions for the 'equal' instruction. NOTE: Dual-task driving performance on the straight route was divided by baseline performance on the straight route, and dual-task driving performance on the curvy route was divided by baseline performance on the curvy route. Str = Straight, Cur = Curvy, Cl = Clean, Dis = Distorted. One-tailed p values were used, except for the transition between 'curvy/clean' and 'straight/distorted'. Significant results in bold, * $p < .05$, ** $p < .01$.

Transition			**Δ Driving performance**			**Δ Memory performance**			**Δ Mental effort**		
From	**To**	**df**	***M***	**t**	**r**	***M***	**t**	**r**	***M***	**t**	**r**
Str_Cl	Str_Dis	22	-.0097	.78	.16	**-17.71**	**2.99****	**.54**	**21.37**	**2.71***	**.50**
Cur_Cl	Cur_Dis	22	-.038 [a]	1.84	.37	-9.37	1.26	.26	7.71	1.09	.23
Str_Cl	Cur_Cl	11	.037	-1.74	.14	**-13.54**	**2.23***	**.16**	**14.11**	**-2.49***	**.30**
Str_Dis	Cur_Dis	11	.0081	-.63	.47	-5.21	.99	.39	.45	-.16	.88
Str_Cl	Cur_Dis	22	-.0016	.11	.023	**-22.92**	**2.97****	**.54**	**21.82**	**-2.76****	**.51**
Cur_Cl	Str_Dis	22	**-.047**	**2.41***	**.46**	-4.17	.68	.14	7.27	.31	.22

[a]: On driving performance the transition between 'curvy/clean' and 'curvy/distorted' was significant ($p < .05$) in the opposite direction of the hypothesized direction.

Table 4.5. *t*-test results for transitions from single-task baselines to dual-task conditions for the 'driving' instruction. *df* = 11. Str = Straight, Cur = Curvy, Cl = Clean, Dis = Distorted. Significant results in bold, * *p* < .05, ** *p* < .01, *** *p* < .001.

Transition			**Δ Driving performance**			**Δ Memory performance**		
From	**To**	***df***	***M***	***t***	***r***	***M***	***t***	***r***
Baseline	Str_Cl	11	.39	.61	.007	**-15.63**	**2.79****	**.18**
Baseline	Cur_Cl	11	-.63	.94	.67	**-16.67**	**3.60****	**.52**
Baseline	Str_Dis	11	-.031	.064	.89	**-22.92**	**3.12****	**.41**
Baseline	Cur_Dis	11	-.22	.55	.86	**-19.79**	**2.29***	**.12**
			Δ Mental effort (baseline: driving)			**Δ Mental effort (baseline: memory)**		
			M	***t***	***r***	***M***	***t***	***r***
Baseline	Str_Cl	11	**22.90**	**5.72*****	**.78**	**-11.00**	**2.53***	**.44**
Baseline	Cur_Cl	11	10.87	1.73	.39	-8.37	1.30	.11
Baseline	Str_Dis	11	**18.79**	**4.13****	**.71**	-13.14	1.74	.11
Baseline	Cur_Dis	11	**16.42**	**2.89****	**.55**	-1.08	.18	.31

Table 4.6. *t*-test results for transitions from single-task baselines to dual-task conditions for the 'equal' instruction. *df* = 11. Str = Straight, Cur = Curvy, Cl = Clean, Dis = Distorted. Significant results in bold, * *p* < .05, ** *p* < .01, *** *p* < .001.

Transition			**Δ Driving performance**			**Δ Memory performance**		
From	**To**	***df***	***M***	***t***	***r***	***M***	***t***	***r***
Baseline	Str_Cl	11	.12	.28	.39	5.21	.86	.28
Baseline	Cur_Cl	11	**1.63**	**2.09***	**.92**	-8.33	1.58	.41
Baseline	Str_Dis	11	-.33	.88	.79	-11.46	1.49	.12
Baseline	Cur_Dis	11	.050	.12	.86	**-16.67**	**2.10***	**.26**
			Δ Mental effort (baseline: driving)			**Δ Mental effort (baseline: memory)**		
			M	***t***	***r***	***M***	***t***	***r***
Baseline	Str_Cl	11	**31.97**	**5.58*****	**.53**	-1.93	.27	.17
Baseline	Cur_Cl	11	**31.42**	**6.25*****	**.53**	**12.18**	**2.50***	**.23**
Baseline	Str_Dis	11	**39.60**	**8.10*****	**.61**	7.67	1.50	.55
Baseline	Cur_Dis	11	**25.61**	**7.24*****	**.84**	8.11	1.40	.42

the memory task. This distinction is relevant, because the addition of the memory task to the driving task (i.e., driving task as baseline for mental effort) may have caused a greater increase in mental effort than the addition of the driving task to the memory task (i.e., memory task as baseline for mental effort). Tables 4.5 and 4.6 display the results of one-tailed t-tests for the 'driving' and 'equal' instructions, respectively. The driving instruction resulted in significantly decreased memory performance on all transitions, with effect sizes ranging from small (i.e., $r > .10$ to large (i.e., $r > .50$). Furthermore, mental effort increased significantly on almost all transitions when the driving task was used as baseline for mental effort, regardless of instruction. These effects were all accompanied by large (i.e., $r > .50$) effect sizes. The significant results in Tables 4.3 to 4.6 have been used to infer coping strategies in the next section.

4.3.4 Coping strategies

So far we have described measurable behavior, in terms of driving performance, memory performance, and mental effort. According to Formula 4.2, changes in these behavioral measures can be used to infer coping strategies, which themselves are not directly measurable. Table 4.1 shows that the direction of change determines which coping strategy, if any, applies. Therefore, we compared the transitions that featured significant effects in Tables 4.3 to 4.6 with the logical combinations of Table 4.1. For example, the transition from 'curvy/clean' to 'curvy/distorted' with the 'driving' instruction did not significantly affect dual-task performance, but it did result in a significant increase in mental effort (see Table 4.3). According to Table 4.1, a 'compensatory costs' strategy can be inferred from this tradeoff. Figure 4.3 displays the resulting coping strategies separately for the 'driving' and the 'equal' instructions.

Clearly, the instructions have resulted in distinct coping behavior, as none of the coping strategies with the 'driving' instruction are found in the 'equal' instruction, and vice versa. The 'compensatory costs' strategy is the only exception, but note that this strategy occurs only at single-to-dual task

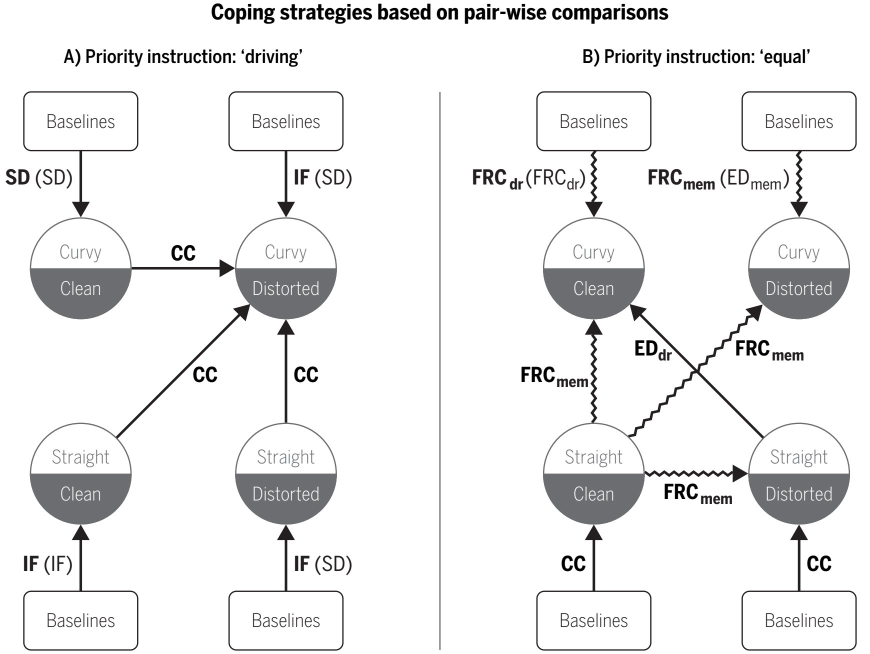

Figure 4.3. Coping strategies inferred from significant tradeoffs between performance and mental effort, given a 'driving' instruction (panel A) or an 'equal' instruction (panel B). 'CC' = Compensatory Costs, 'ED' = Exclusive Decrement, 'IF' = Intense Focus, 'SD' = Secondary Decrement, 'dr' = driving performance decrement, 'mem' = memory performance decrement. Note that Failed Regulatory Control ('FRC', zigzag lines) is not regarded as coping strategy. Single driving task conditions were used as baseline for mental effort (parentheses: single memory task condition as baseline, see text for details).

transitions with 'driving' instruction, and only at dual-to-dual task transitions with the 'equal' instruction. Furthermore, the transition from 'straight/clean' to 'curvy/distorted' is the only significant transition found with both instructions.

If the source of increased task demand influences coping behavior, then within each priority instruction the manipulation from 'straight' to 'curvy' should result in different coping strategies than the manipulation from 'clean' to 'distorted'. The symmetrical distribution of dual-to-dual task transitions in Figure 4.3 show that this is not the case. The only exception is the 'exclusive decrement' strategy from 'straight/distorted' to 'curvy/clean', with decreased performance on the driving task. Note that two sources of task demand were simultaneously and oppositely manipulated in this transition. Therefore, we do not involve the resulting strategy when we evaluate the suggestion of Cnossen et al. (2004).

Contrary to what was predicted, all dual-to-dual task transitions with the ‘driving’ instruction resulted in a ‘compensatory costs’ strategy (see Figure 4.3A). Instead, the predicted ‘secondary decrement’ strategy occurred at some of the transitions from single-task to dual-task conditions. The remaining transitions featured the newly proposed ‘intense focus’ strategy. Another unexpected finding is that no ‘compensatory costs’ strategy was found in the dual-to-dual task transitions with the ‘equal’ instruction (see Figure 4.3B). This strategy instead occurred at single-to-dual task transitions with clean news items. Finally, five transitions with the ‘equal’ instruction resulted in increased mental effort, combined with either decreased memory or driving performance. These tradeoffs were interpreted as ‘failed regulatory control’.

4.4 Discussion

This study presented the logics to infer coping strategies from tradeoffs between task performance and effort, with explicit attention to task prioritization. The logics were applied in a dual-task experiment with a driving task and a memory task. Our main finding is that coping strategies were selected as function of task prioritization, but contrary to our prediction, not as function of source of increased task demand.

4.4.1 Source of increased task demand

Within each priority instruction, the manipulation of driving task demand did not result in different coping strategies than the manipulation of memory task demand. This finding opposes the suggestion of Cnossen et al. (2004) that the source of increased task demand influences the selection of coping strategies. The difference between our findings and those by Cnossen et al. (2004) may be explained by a study by Alm and Nilsson (1994), who suggested that drivers change their self-chosen task prioritization depending on road curvature. In their study, the addition of an auditory memory task to a driving task resulted in lower driving speed on a route with straight roads (i.e., increased priority on the memory task), but not on a route with curvy roads (i.e., a 'driving' task prioritization). A similar mechanism may have affected the study of Cnossen et al. (2004), who also did not provide priority instructions. They inferred from their results that the driving task was protected (i.e., a supposed 'driving' task prioritization). Participants drove faster in quiet traffic than in busy traffic, but in quiet traffic more navigation errors were observed. Thus, it appears that those participants shifted their priority towards the memory task when driving task demand decreased (cf. Alm & Nilsson, 1994). We have shown how coping strategies should be inferred based on increased task demand, and importantly, given a fixed task prioritization. If, as we suggest above, task prioritization was unstable in the study by Cnossen et al. (2004), then the suggestion that coping behavior depends on the source of increased task demand may have been confounded by momentary task prioritization. This demonstrates the importance of systematically manipulating and verifying task prioritization, as was done in the present study.

A disclaimer is at place, though, in that there are two reasons why the likeliness of finding decreased driving performance and decreased memory performance may not have been equal. First, differences in effort ratings indicate that the demands of the memory task were higher than those of the driving task. Second, and potentially influencing the former reason, the driving task was practiced until stable single-task performance arose (possibly accompanied by relatively low mental effort, cf. Charlton & Starkey, 2011), whereas the memory task did not receive such extensive practice. Additional single-task practice on the memory task may have improved performance and effort (Brookhuis et al., 1991), although we note that no such effect was found with a single-task control group in our previous study (Jansen et al., 2016). To summarize, while our study does not support Cnossen et al.'s (2004) suggestion, neither does our study provide the means to refute it.

4.4.2 Coping strategies resulting from task prioritization

Through logical reasoning we have shown that two coping strategies may exist in addition to the ones predicted by Hockey's (1997, 2011) Compensatory Control Model, namely 'exclusive decrement' and 'intense focus'. Our experiment provided empirical evidence for the existence of both strategies. Furthermore, our experiment demonstrated the role of priority instructions on coping behavior. In line with our previous work (Jansen et al., 2016), memory performance generally proved to be higher with the 'equal' instruction than with the 'driving' instruction, but at the cost of increased mental effort. However, none of the predicted coping strategies were found in transitions between dual-task conditions.

The 'driving' instruction yielded the 'compensatory costs' strategy, where we had expected to find either 'strategic adjustment', or 'secondary decrement'. Extensive single-task practice on the driving task (see above) may explain the absence of the 'strategic adjustment' strategy. Figure 2.B shows that memory performance per signal type was higher in single-task baselines than in the corresponding dual-task conditions. Thus, the expected 'secondary decrement' coping strategy appeared to have occurred in the transition from single-task to dual-task, and not in transitions between dual-task conditions. Further increases in dual-task demands were then countered by investing more effort, resulting in the 'compensatory costs' strategy.

Moving on to the 'equal' instruction, we had expected to find the 'compensatory costs' strategy, but instead the instruction mostly resulted in 'failed regulatory control'. In all of these cases, increased effort was accompanied by decreased memory performance, but not by decreased driving performance. Two mechanisms may explain these findings, namely preferences and overload. In our previous study (Jansen et al., 2016) we recruited participants from the same population and used similar tasks as in the present study. Participants with a 'driving' preference (i.e., a preference to prioritize the driving task) consistently showed lower memory performance than participants with an 'equal' preference, even though the ‘equal’ instruction was in fact followed by both preference groups. Furthermore, we previously found that there were more participants with a ‘driving’ preference than an ‘equal’ preference. Given the similarities with the present study, preference in task prioritization may explain why it was the memory task, and not the driving task, that suffered from decreased performance with the ‘equal’ instruction.

Regarding overload, mental effort ratings between 65 and 90 (i.e., ‘considerable effort’ to ‘great effort’) suggest that the memory task demands may have been too high. It should be noted that compared to Cnossen et al. (2004), who reported effort ratings between 40 and 60, our stimuli had a longer duration, and questions were not known in advance. It appears as if participants were initially able to cope with increased task demands through a ‘compensatory costs’ strategy, that is, from single-to-dual task conditions with the straight route. But when the dual-task demands increased even further, it may no longer have been possible to invest more effort. Consequently, memory performance decreased, which we interpreted as ‘failed regulatory control’. The absence of this failure at the transitions from ‘curvy/clean’ and ‘straight/distorted’ to ‘curvy/distorted’ could in turn be interpreted as a ceiling effect (i.e., all rated at or around ‘great effort’). To summarize, it appears that there is an upper limit to task demands, beyond which regulatory control is no longer possible.

4.4.3 Implications

These findings have several implications for future research on the selection of coping strategies. From a theoretical perspective, knowledge on task prioritization is essential to infer coping strategies from tradeoffs between dual-task performance and mental effort. Furthermore, Hockey’s (1997, 2011) Compensatory Control Model should be extended to include the ‘exclusive decrement’ and ‘intense focus’ coping strategies. This way it is possible to infer coping strategies from all logical combinations of dual-task performance and effort, rather than a subset of these combinations.

From a methodological perspective, the occurrence of failed regulatory control with the ‘equal’ instruction raises the question which task demand levels allow for the selection of coping strategies. Coping behavior is initiated by an increase in task demands (Hockey, 1997), where a minimum increase is required to detect coping in terms of tradeoffs between dual-task performance and mental effort. Our findings support an additional qualifier on detecting coping behavior: when task demand reaches an upper limit, a further increase in task demand will not yield significant effects on performance and mental effort. Therefore, future research should control not only for task prioritization, but also for an upper limit in task demand, when testing Cnossen et al.’s (2004) suggestion that coping behavior is guided by the source of increased task demand.

Finally, we return to the case of the police officers. Our previous observation study (Jansen et al., 2012) showed that officers are not always able to follow incoming portophone messages while driving, despite investing effort to do so. In line with that observation study, we found ‘failure in regulatory control’ with the ‘equal’ instruction (i.e., decreased memory performance and increased effort), the latter of which was given to emulate police officers’ task prioritization. On the other hand, the ‘driving’ instruction, which was intended to emulate regular drivers’ task prioritization, did not result in ‘failed regulatory control’. This finding implies that situational task demands are often too high for police officers to employ successful coping strategies, whereas regular drivers do have the means to compensate for such demands. Given that police officers have no control over incoming radio messages, and given that they do not experience successful performance, it is not surprising to have observed fatigue-related complaints (Jansen et al., 2012; Hockey, 2011). Therefore, the present study strongly supports a reconsideration of current radio communication practice in police work.

Acknowledgments

This research was partially funded by Staf Korpsleiding / Directie Facility Management, Nationale Politie.

References

Aarts, L. & van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention 38*(2): 215-224.

Acheson, D.J. & MacDonald, M.C. (2009). Verbal Working Memory and Language Production: Common Approaches to the Serial Ordering of Verbal Information. *Psychological Bulletin 135*(1): 50-68.

Alm, H.K. & Nilsson, L. (1994). Changes in driver behaviour as a function of handsfree mobile phones: A simulator study. *Accident Analysis & Prevention 26*(4): 441-451.

Anderson, G.S., Courtney, A., Plecas, D. & Chamberlin, C. (2005). Multi-tasking behaviors of general duty police officers. *Police Practice and Research 6*(1): 39-48.

Baldwin, C.L. (2007). Cognitive implications of facilitating echoic persistence. *Memory & Cognition 35*(4): 774-780.

Boersma, P. & Weenink, D. (2013). *Praat: doing phonetics by computer*. Computer program, version 5.3.51, retrieved 2 June 2013 from http://www.praat.org

Brookhuis, K.A., de Vries, G. & de Waard, D. (1991). The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention 23*(4): 309-316.

Caird, J.K., Willness, C.R., Steel, P. & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention 40*(4): 1282-1293.

Charlton, S.G. & Starkey, N.J. (2011). Driving without awareness: The effects of practice and automaticity on attention and driving.

Transportation Research Part F: Traffic Psychology and Behaviour 14(6): 456-471.

Clarke, D.D., Ward, P., Bartle, C. & Truman, W. (2009). Work-related road traffic collisions in the UK. *Accident Analysis & Prevention 41*(2): 345-351.

Cnossen, F., Meijman, T. & Rothengatter, T. (2004). Adaptive strategy changes as a function of task demands: a study of car drivers. *Ergonomics 47*(2): 218-236.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Routledge.

Dingus, T.A., Guo, F., Lee, S., Antin, J.F., Perez, M., Buchanan-King, M. & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *PNAS 113*(10): 2636-2641.

Dressel, J. & Atchley, P. (2008). Cellular phone use while driving: A methodological checklist for investigating dual-task costs. *Transportation Research Part F: Traffic Psychology and Behaviour 11*(5): 347-361.

Gopher, D. & Donchin, E. (1986). Workload - An examination of the concept. In: Boff, K.R., Kaufman, L. & Thomas, J.P. (Eds.) *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance*. Oxford, England, John Wiley & Sons.

Gopher, D. & Navon, D. (1980). How is performance limited: Testing the notion of central capacity. *Acta Psychologica 46*(3): 161-180.

Haigney, D.E., Taylor, R.G. & Westerman, S.J. (2000). Concurrent mobile (cellular) phone use and driving performance: task demand characteristics and compensatory processes. *Transportation Research Part F: Traffic Psychology and Behaviour 3*(3): 113-121.

Hancock, P.A., Mouloua, M. & Senders, J.W. (2008). On the Philosophical Foundations of the Distracted Driver and Driving Distraction. *Driver Distraction*, CRC Press: 11-30.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology 45*(1-3): 73-93.

Hockey, G.R.J. (2011). A motivational control theory of cognitive fatigue. In: Ackerman, P.L. (Ed.) *Cognitive fatigue: multidisciplinary perspectives on current research and future applications*. Washington, DC, American Psychological Association: 167-188.

Hockey, G.R.J. & Earle, F. (2006). Control over the scheduling of simulated office work reduces the impact of workload on mental fatigue and task performance. *Journal of Experimental Psychology: Applied 12*(1): 50-65.

Horberry, T., Anderson, J., Regan, M.A., Triggs, T.J. & Brown, J. (2006). Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention 38*(1): 185-191.

Jamson, A.H. & Merat, N. (2005). Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. *Transportation Research Part F: Traffic Psychology and Behaviour 8*(2): 79-86.

Jansen, R.J., Van Egmond, R., De Ridder, H. & Silvester, S. (2014). Transitional Journey Maps: Capturing the dynamics of operational policing. In: De Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W. & Toffetti, A. (Eds.) *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*, 16-18 October 2013, Turin, Italy.

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2016). Task Prioritization in Dual-Tasking: Instructions versus Preferences. *PLoS ONE 11*(7): e0158511.

Janssen, C.P., Brumby, D.P. & Garnett, R. (2012). Natural Break Points: The Influence of Priorities and Cognitive and Motor Cues on Dual-Task Interleaving. *Journal of Cognitive Engineering and Decision Making 6*(1): 5-29.

Jordan, P.W. & Johnson, G.I. (1993). Exploring mental workload via TLX: the case of operating a car stereo whilst driving. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Kruysse, H.W. & Taylor, S.P. (Eds.) *Vision in Vehicles IV*. Amsterdam, North-Holland: 255-262.

Lee, J.D., Caven, B., Haake, S. & Brown, T.L. (2001). Speech-Based Interaction with In-Vehicle Computers: The Effect of Speech-Based E-Mail on Drivers' Attention to the Roadway. *Human Factors 43*(4): 631-640.

Liepelt, R., Strobach, T., Frensch, P. & Schubert, T. (2011). Improved intertask coordination after extensive dual-task practice. *The Quarterly Journal of Experimental Psychology 64*(7): 1251-1272.

Norman, D.A. & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology 7*(1): 44-64.

O'Donnell, R.D. & Eggemeier, F.T. (1986). Workload assessment methodology. In: Boff, K.R., Kaufman, L. & Thomas, J.P. (Eds.) *Cognitive Processes and Performance*. John Wiley and Sons.

Rakauskas, M.E., Gugerty, L.J. & Ward, N.J. (2004). Effects of naturalistic cell phone conversations on driving performance. *Journal of Safety Research 35*(4): 453-464.

Schultz (2012). *RC Mini Racers*. Producer: Decane. Retrieved December 1, 2012, from http://www.decane.net/game/mac-app-store/rc-mini-racers/

Sørensen, C. & Pica, D. (2005). Tales from the police: Rhythms of interaction with mobile technologies. *Information and Organization 15*(2): 125-149.

Strayer, D.L., Drews, F.A. & Johnston, W.A. (2003). Cell Phone-Induced Failures of Visual Attention During Simulated Driving. *Journal of Experimental Psychology: Applied 9*(1): 23-32.

Törnros, J.E.B. & Bolling, A.K. (2005). Mobile phone use - Effects of handheld and handsfree phones on driving performance. *Accident Analysis & Prevention 37*(5): 902-909.

Tsang, P.S., Velazquez, V.L. & Vidulich, M.A. (1996). Viability of resource theories in explaining time-sharing performance. *Acta Psychologica 91*(2): 175-206.

Ünal, A.B., Platteel, S., Steg, L. & Epstude, K. (2013). Blocking-out auditory distracters while driving: A cognitive strategy to reduce task-demands on the road. *Accident Analysis & Prevention 50*: 934-942.

Zar, J.H. (1996) *Biostatistical Analysis*, 3rd ed., p282-283. Upper Saddle River, New Jersey: Prentice Hall.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior - A design approach for modern tools*. Delft, The Netherlands, Delft University of Technology. PhD thesis.

Chapter 5 has been published as:

Jansen, R.J., Sawyer, B.D., van Egmond, R., de Ridder, H. & Hancock, P.A. (2016). "Hysteresis in Mental Workload and Task Performance: The Influence of Demand Transitions and Task Prioritization." Human Factors 58(8): 1143-1157.

Chapter 5

Hysteresis in mental workload and task performance: the influence of demand transitions and task prioritization

Abstract

Objective: This study examines how transitions in task demand are manifested in mental workload and performance, in a dual-task setting. **Background:** Hysteresis has been defined as the on-going influence of demand levels prior to a demand transition. Previous studies predominantly examined hysteretic effects in terms of performance. However, little is known about the temporal development of hysteresis in mental workload. **Method:** A simulated driving task was combined with an auditory memory task. Participants were instructed to prioritize driving, or to prioritize both tasks equally. Three experimental conditions with low, high, and low task demands were constructed by manipulating the frequency of lane changing. Multiple measures of subjective mental workload were taken during experimental conditions. **Results:** Contrary to our prediction, no hysteretic effects were found after the high-to-low demand transition. However, an hysteretic effect in mental workload was found within the high demand condition, which degraded toward the end of the high condition. Priority instructions were not reflected in performance. **Conclusion:** Online assessment of both performance and mental workload demonstrates the transient nature of hysteretic effects. An explanation for the observed hysteretic effect in mental workload is offered in terms of effort regulation. **Application:** An informed arrival at the scene is important in safety operations, but peaks in mental workload should be avoided to prevent buildup of fatigue. Therefore, communication technologies should incorporate the historical profile of task demand.

5.1 Introduction

> *After a short break at the police station, C. and S. return to their surveillance duty. The dispatcher calls: 'A missing girl possibly showed up at relatives and should be picked up.' While S. tries to write down the address in his notebook, they realize they missed the girl's full name and the house number. S. feels stupid for having to ask again. Directly afterwards an alarm goes off. S. glances at the mobile data terminal: 'It's a white vehicle with an unpaid fine.' C. looks around, locates the car, and immediately makes a turn. Just as S. tries to request information on the driver, the dispatcher interrupts him: 'We are detaching you from the previous call. Someone has been spotted in a building that burned down last week.' C. recognizes the address, turns the car again, and accelerates. On their way, S. declines another alarm with a lower priority. They arrive at the scene only minutes later, to find a man in ragged clothes carrying a bag full of copper.* (field notes from Jansen et al., 2014)

This anecdote illustrates how police officers continuously perform in-vehicle tasks while driving, such as memorizing incoming radio messages, verbal communication, and operating the mobile data terminal (Anderson et al., 2005; Jansen et al., 2014). It also makes clear that, unlike regular traffic participants, police officers do not have the choice to ignore incoming messages. Within this multi-task context, police officers are engaged in a variety of activities with different levels of task demand. For example, rushing to catch a thief likely imposes a greater driving task demand than transporting said thief to the police station, because the former activity requires driving at a higher speed while avoiding other traffic. Police work is also characterized by a perpetual switching between activities (Borglund & Nuldén, 2012; Jansen et al., 2014; Sørensen & Pica, 2005), leading to frequent and sudden transitions between high and low task demands.

The absolute demand level prior to a sudden demand transition is known to affect performance and mental workload for a certain period of time directly after such transition occurs (for overviews, see: Cox-Fuenzalida, 2007; Morgan & Hancock, 2011). This ongoing influence of prior demand level is referred to as 'hysteresis' (Cumming & Croft, 1973; Farrell, 1999; Goldberg & Stewart, 1980; Morgan & Hancock, 2011). Previous studies have shown that hysteresis degrades over time in terms of performance (Gluckman et al., 1993; Matthews, 1986). However, surprisingly little is known about how this temporal nature of hysteresis affects concomitant mental workload.

Our present study focuses on how hysteresis after demand transitions evolves over time, both in terms of performance and mental workload. In line with the above police example, a driving task was combined with an auditory memory task. Demand transitions were induced by manipulating the difficulty of the driving task, while keeping the auditory memory task at the same demand level. In addition, task prioritization was manipulated between participant groups to reflect the constraint that police officers do not have the choice to ignore incoming messages, whereas regular drivers do. This study informs an understanding of existing theories on hysteresis (i.e., resource depletion, effort regulation) by assessing mental workload not only after, but also during on-going experimental performance.

5.1.1 Hysteresis in performance

Several studies have described how hysteresis develops over time by comparing multiple periods of aggregated performance data (Matthews, 1986; Gluckman et al., 1993; Ungar et al., 2005; Cox-Fuenzalida, 2007). Matthews (1986) aggregated performance on a visual signal-detection task in fifteen consecutive periods of 10 seconds each. Task demand was manipulated by varying the number of co-occurring stimuli. A sudden transition from high to low task demand resulted in an immediate performance reduction that lasted for six periods (i.e., 1 min in total), before returning to a performance level similar to that of a low task demand control group. Gluckman et al. (1993) aggregated performance over a longer period of time, but at a lower temporal resolution. Demand transitions were induced by shifting from two parallel visual signal-detection tasks to one signal-detection task, or vice versa. Pre-transition and post-transition performance were both measured in two periods of 10 minutes, and compared against non-shifting control groups. A hysteresis effect in the form of lower performance was found only with the shift from dual-task to single-task. This effect was found in the first period of 10 minutes after the demand transition, but not in the second period.

The above studies demonstrate that hysteresis in performance decays over time. Two other studies, however, indicate that sudden demand transitions can have permanent hysteretic effects. Ungar et al. (2005) induced a demand transition by shifting from a compensatory tracking task with a visual signal-detection task, to the compensatory tracking task only. The resulting hysteretic effect persisted throughout all post-transition periods (i.e., 8 x 2 min). Cox-Fuenzalida (2007) manipulated the difficulty of an auditory signal-detection task, and also found a hysteretic effect that persisted throughout all post-transition periods (i.e., 3 x 3 min). Although in these studies hysteresis seems to be permanent, it should be noted that different experimental tasks and modalities were used, as compared to those by Matthews (1986) and Gluckman et al. (1993). It is still possible that the experimental conditions in Ungar et al. (2005) and Cox-Fuenzalida (2007) were too short to measure any existing decay in hysteresis.

Regardless, these studies show that the partitioning of performance data into a sequence of post-transition periods is essential to investigate how hysteresis develops during the time frame of an experimental condition, and thus potentially in real-world situations.

Gluckman et al. (1993) interpreted the finite duration of hysteresis, and the fact that hysteresis only occurred at a transition to lower demands, in terms of two theories: ‘resource depletion’ and ‘effort regulation’. The resource depletion theory is an analogy to the recovery of muscle tissue in exercise physiology (Cannon, 1932). The high demands of dual-tasking may have caused a resource debt, which could be satisfied through temporary regeneration after a transition to a lower demand. Alternatively, the findings were explained in terms of an effort regulation theory (see Hancock & Warm, 1989), in which increased mental effort is viewed as a means of regulating resources under varying demands (Hockey, 1997). Initial dual-tasking may have formed a policy to distribute resources across the two interfering tasks. If this resource allocation policy was maintained after the transition to the single-task, then the remaining task receives suboptimal resource allocation. Continued single-task exposure led to a revision of policy. In other words, the resource depletion theory interprets hysteresis in terms of recuperation, whereas the effort regulation theory interprets hysteresis in terms of strategic persistence. A next question, then, is whether hysteresis also degrades over time in terms of mental workload.

5.1.2 Hysteresis in mental workload

Three studies on demand transitions have assessed mental workload in addition to performance. In the first of these, Hancock et al. (1995) subjected participants to three trials on a compensatory tracking task. The first and the third trial were performed at an identical difficulty level. When the second trial was set to a lower difficulty level, mental workload (i.e., NASA-TLX and SWAT ratings) in the third trial increased compared to the first trial. Conversely, when the second trial was set to a higher difficulty level, mental workload in the third trial was rated lower than the first trial. Matthews and Desmond (2002) induced a demand transition by shifting from a dual-task driving trial (i.e., with a signal-detection task) to a single-task driving trial. Shifted drivers reported higher mental workload on the NASA-TLX than non-shifted drivers. Furthermore, shifted drivers showed impaired driving performance on straight road sections in the single task, but not on curvilinear road sections. Morgan and Hancock (2011) also used a driving setting. Task demand was temporarily increased with a problem-solving task halfway through the drive. The problem-solving task increased mental workload on the S-SWAT compared to a baseline measure, and this increase persisted until the end of the drive. The mental effort component of the S-SWAT proved to be the only contributor to this hysteresis. The effect was attributed to short-term memory overload (cf. Reid & Nygren, 1988), akin to the resource depletion theory (Gluckman et al., 1993).

The above three studies demonstrate that hysteresis is also manifested in mental workload. However, none of them partitioned the data in a sequence of post-transition periods, since subjective workload ratings were only collected once after each experimental condition. Consequently, the development of mental workload during experimental conditions could not be investigated. Moreover, the hysteretic effects may have lasted longer than the duration of an experimental condition (i.c., 2 to 5 min).

Frequent online ratings of subjective mental workload appear to solve the above problem. One concern with online ratings, however, is that they may be intrusive to the experimental tasks. A study by Hill et al. (1992) suggested that intrusiveness can be minimized with uni-dimensional rating scales, as opposed to multi-dimensional rating scales (e.g., NASA-TLX, SWAT). However, this strategy comes at the expense of reduced diagnosticity. Morgan and Hancock (2011) showed why high diagnosticity is important for the interpretation of hysteresis in mental workload. An appropriate balance between low intrusiveness and high diagnosticity may be obtained by combining an online uni-dimensional scale with a multi-dimensional scale at the end of each experimental condition, where the latter scale is used to interpret the former. Such an effort is thus enacted here.

5.1.3 Paradigm

The present study examines hysteresis in a dual-task setting with a continuous driving task and a continuous auditory memory task, inspired by operational policing (Jansen et al., 2014). Driving task demand was manipulated by increasing the frequency of lane changing maneuvers. This manipulation resulted in three experimental dual-task conditions with a low-high-low demand schedule. Hysteretic effects were tested by comparing driving performance, memory performance, and mental workload across the conditions with low demands (cf. Hancock et al., 1995).

The unidimensional Instantaneous Self Assessment (ISA) scale (Jordan, 1992; Tattersall & Foord, 1996; Leggatt, 2005) was used to assess mental workload during experimental conditions. The ISA scale was initially developed for the aviation context, but recent studies have shown that it is also sensitive to variations in traffic conditions (Girard et al., 2005), and to the distraction of a memory task while driving (Lemercier et al., 2014). The original ISA protocol uses a visual signal to prompt participants to rate their experienced workload level on a keypad. To minimize interference with the visual/manual driving task, we adapted this protocol by using an auditory trigger and by eliciting verbal, numeric responses. ISA prompts did not co-occur with auditory memory items to minimize interference with the memory task. In addition to the ISA ratings, the NASA-TLX (Hart & Staveland, 1988) was administered at the end of each experimental condition. NASA-TLX sub-scales (i.e., mental demand, physical demand, temporal demand, subjective performance, effort, frustration) were used to interpret ISA ratings collected during the experimental conditions.

Although radio communication is viewed as a distraction and a risk for all drivers (Caird et al., 2008; Dressel & Atchley, 2008), police officers do not have the choice to ignore incoming messages. As a result, police officers frequently have to de-prioritize the driving task in favor of secondary tasks, especially in case of solo patrols. We addressed this difference in task prioritization through separate instructions for solo patrols (i.e., an ‘equal’ instruction) and regular driving (i.e., a ‘driving’ instruction).

The manipulation of task prioritization could provide an alternative paradigm to test the theories proposed by Gluckman et al. (1993). The effort regulation theory can be tested with the ‘driving’ instruction, which should result in a protection of driving at the cost of memory performance. Thus, memory performance should decrease as the driving task becomes more demanding. The effort regulation theory predicts a temporary persistence of this resource allocation policy. A sudden decrease in driving demand should then result in impaired memory performance, but not in impaired driving performance. The resource depletion theory, on the other hand, can be tested through a comparison between the instructions. The ‘equal’ instruction is likely to confront drivers with higher overall demands than the ‘driving’ instruction (Jansen et al., 2016; Kantowitz & Knight, 1976; Norman & Bobrow, 1975). The higher the resulting resource debt, the longer one can expect recuperation to last after a high-to-low demand transition. Therefore, the resource depletion theory predicts a longer hysteretic effect with the ‘equal’ instruction set.

5.2 Method

5.2.1 Participants

Twenty-eight students from the Department of Psychology at the University of Central Florida (UCF) were recruited to participate in the experiment. There were 10 males and 18 females, ranging in age from 18 to 41 years old, (*M* = 19.9 years, *SD* = 4.4). They were compensated for their time with class credit. The study was approved by UCF’s ethical committee. Participants had normal or corrected-to-normal vision. All participants had a current driver’s license (*M* = 3.4 years, *SD* = 4.0), and on average they drove 179 km (111 miles) per week (*SD* = 205 km, 127 miles).

5.2.2 Auditory memory task

An initial set of 98 auditory stimuli was compiled, consisting of American radio news items spoken by professional newsreaders (‘Here&Now’, http://hereandnow.wbur.org/section/radio). Selected news items were at least 1.5 years old to minimize recency effects. A native speaker from the United States recorded a factual question for each news item. Questions were related to numbers or names close to the center of each news item, and allowed for only one correct answer. For example, the item: *“Mixing chemicals in a high school lab is challenging enough. Now imagine you are doing it blind. A group of visually impaired students from all over the country had that chance at Metro State University in Denver recently, as part of an effort to get more blind people interested in science, technology, and math.”* was accompanied by the question: *“In which city was the university located?”* The news items and questions were normalized to the same average volume, and saved as wav files (16 bit, 44.1 kHz).

In an initial pilot procedure with 15 participants, each item plus its accompanying question was presented to five participants. Questions that were answered incorrectly by more than 75% of these participants were removed. The final auditory stimulus set consisted of 64 news items (M = 17.2 sec, SD = 1.20 sec). Sixteen of these were used for training, whereas the other 48 were used in three experimental conditions. The goal of the memory task was to correctly answer a question for each stimulus.

5.2.3 Driving task

The goal of the driving task was to avoid obstacles on a roadway, without speeding. A simulated driving environment was created by placing obstacles on a straight 8.5 km (5.3 miles) section of a simulated three-lane freeway. No additional traffic was added to these scenarios. The straight section could not be finished within the duration of any one experimental condition, if the driver followed the posted speed limit of 56 km/h (35 mph).

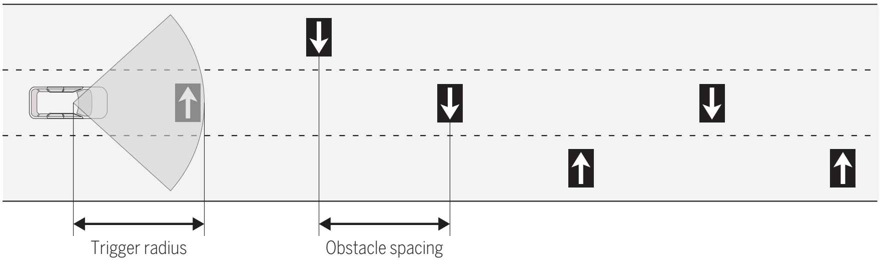

Figure 5.1. Track layout with obstacles. The displayed ratio between trigger radius and obstacle spacing corresponds with the ‘easy’ obstacle map.

In terms of spatial distribution, the first obstacle and ensuing odd obstacles were positioned on the center lane (see Figure 5.1). The other interpolated obstacles were pseudo-randomly distributed in the two outer lanes. The obstacles were composed of flashing arrow signs that required the driver to change lane. Arrows pointed rightwards when positioned on the left lane, and vice versa. In the center lane the arrows pointed in the direction of the next obstacle (i.e., either left or right).

The first obstacle was positioned at 0.3 km (0.19 miles) into the drive. Subsequent obstacles were equally distributed over the remaining freeway section. Three obstacle maps were created, which differed in the longitudinal distribution of the obstacles. $\mathrm{MAP}_{\mathrm{training}}$ had an obstacle spacing of 100 m (328 ft). In $\mathrm{MAP}_{\mathrm{far}}$ and $\mathrm{MAP}_{\mathrm{near}}$, the obstacle spacings were 150 m (492 ft) and 82 m (269 ft), respectively. Additionally, the obstacle trigger radius was varied across map conditions. The driver could see two obstacles ahead in $\mathrm{MAP}_{\mathrm{training}}$, due to a trigger radius of 200 m (656 ft). In $\mathrm{MAP}_{\mathrm{far}}$, the next obstacle appeared as the driver passed an obstacle (trigger radius: 148 m, 486 ft), whereas obstacles appeared relatively sudden in $\mathrm{MAP}_{\mathrm{near}}$ (trigger radius: 50 m, 164 ft). Although hitting obstacles did not affect driving speed, the instruction set explicitly required drivers to avoid all obstacles, and to answer all memory questions correctly.

5.2.4 Experimental design and measures

A mixed design was used, with driving task demand as repeated factor and priority instruction as between-subjects factor. The ‘driving’ instruction was to prioritize the driving task over the memory task, whereas the ‘equal’ instruction required the driver to perform both tasks as well as possible. Driving task demands were manipulated through a combination of obstacle spacing and trigger radius. Three experimental conditions were used with a fixed order: LOW_1, $HIGH_2$, and LOW_3 (see Figure 5.2). MAP_{far} and the corresponding trigger radius were used in the LOW_1 and LOW_3 conditions. MAP_{near} was used in the $HIGH_2$ condition. Memory task demand was not changed across driving conditions. Hysteresis was examined by comparing performance and mental workload in LOW_3 with LOW_1.

Two measures of mental workload were taken, namely NASA-TLX and ISA. The NASA-TLX was administered after each condition. Furthermore, the memory task was interleaved with ISA prompts, which followed after each block of four memory trials (see Figure 5.2). ISA mental workload ratings were collected verbally. A pilot study suggested that a 5-point scale was insufficiently sensitive to discern between the high and low levels of driving task difficulty. Therefore, a 7-point scale was used, where ‘1’ corresponded with a ‘very easy task’, and ‘7’ with a ‘very difficult task’.

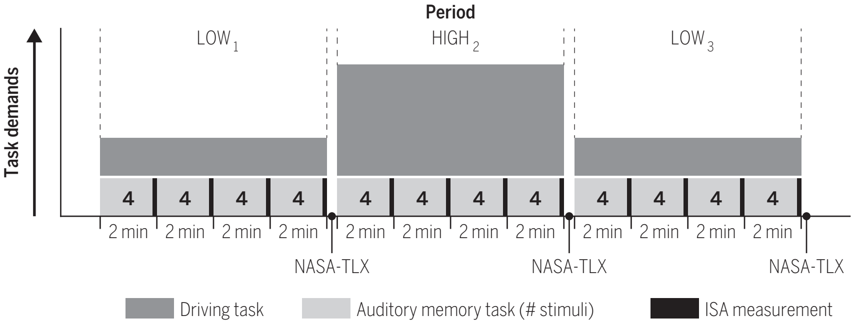

Figure 5.2: Task demands as function of the experimental conditions LOW_1, $HIGH_2$, and LOW_3.

Driving performance and memory performance were calculated over two time frames: per period (8 min. each), and per 2 min. trial block (i.e., four memory trials with an ISA prompt). Driving performance was measured in terms of absolute velocity as well as the root-mean-square-error (RMSE) of the velocity. Memory performance was measured as the proportion of correct answers. These proportions were transformed via an arcsine transformation (Zar, 1996, p.282) for subsequent statistical analyses. All statistical tests were conducted with SPSS v.22, and results were tested using an α level of .05.

5.2.5 Apparatus

A fixed platform police training simulator was used (L3 STS, Inc.). The simulator featured a cab complete with steering wheel and dashboard from an actual automatic transmission vehicle. Three 52 inch screens (1024x768 pixels at 60 Hz) mounted at a distance of approximately 1.0 m from the driver provided a 120 degree view of the driving environment (and see Morgan & Hancock, 2011). Driving speed was sampled at 60 Hz. The NASA-TLX was administered via Qualtrics.com, presented on a tablet next to the simulator. A dedicated program, coded in Max v.6 (Cycling74, Inc.), was used to play pre-recorded instructions, to randomize stimuli for each participant, to collect demographic information, and to record verbal responses to memory trials and ISA prompts through an external microphone. Sounds were played back over a pair of Altec Lansing AVS200 computer speakers, positioned on the dashboard. Auditory stimuli were presented at a comfortable listening level, clearly audible above the simulator sounds. Collection of driving performance measures and trigger behavior were handled through custom software (see Sawyer & Hancock, 2012).

5.2.6 Experimental procedure

Participants were randomly assigned to either the ‘driving’ instruction (n = 14) or the ‘equal’ instruction (n = 14). Each session was organized in three phases: training, experimentation, and interview. Upon arrival, participants were asked to complete the informed consent, and to turn off all electronic devices. Each participant was then trained in the memory task and in responding to ISA trials (i.e., *"Report how much mental workload the task just required."*). Memory trials were composed of a news item (*M* = 17.2 sec, *SD* = 1.2 sec), 1.0 sec silence, a question (*M* = 2.6 sec, *SD* = .6), 4.5 sec answer time, a beep sound, and 2.0 sec silence. After every fourth memory trial, an ISA trial was triggered in order to obtain a mental workload rating. Such trials started with a chicken 'squawk', a salient prompt to attract attention, then provided 7.0 sec answer time, a beep sound, and 2.0 sec silence. Participants were trained to verbalize an answer during the designated interval. Answering after the beep sounds of the memory trials and ISA prompts was permitted, but the participant was urged to prepare for the next trial. Four minutes of single-task practice followed, consisting of eight memory trials, presented in random order, and interleaved by two ISA prompts.

Dual-task training took place in the driving simulator, using the $\text{MAP}_{\text{training}}$ condition. Participants were instructed to shift lanes according to the direction of the arrows, and to maintain driving speed at or below 35 mph (56 km/h). A 'driving' or an 'equal' priority instruction was given, depending on the allocated group. Regardless of the instruction, each participant was directed to provide a rating, related to the combination of both tasks in response to ISA prompts. The first memory trial was triggered directly after the participant had passed the fourth obstacle (i.e., after approximately 1 min). The timing of subsequent memory trials, ISA prompts, and obstacles, was not synchronized. As with the memory training, two ISA prompts took place amidst eight memory trials (i.e., 4 min in total). The remaining training stimuli were used. Upon completion, the participant was instructed to stop and turn off the vehicle. Demographic information was obtained afterwards.

The experimental condition LOW_1 employed the same instructions as dual-task training. However, the first memory trial was already triggered as the second obstacle became visible. Sixteen memory trials and four ISA prompts were presented, with a total duration of eight minutes. The auditory tracks were pseudo-randomly selected from a pool of 48 experimental tracks, such that between participants the stimuli were counterbalanced over the three experimental conditions. Furthermore, each block of four memory trials had a similar distribution of number- and name-related questions. The participant completed a NASA-TLX questionnaire after turning off the vehicle. The experimental conditions $HIGH_2$ and LOW_3 started as soon as the NASA-TLX of the previous condition had been completed. The remaining stimuli were presented according to the above protocol. The priority instruction was repeated before each condition. Sessions ended with open questions about the overall dual-task experience during the experiment, what strategy was used, how it felt to act according to a priority instruction, recognition of news items, and news listening habits.

5.3 Results

Five participants were excluded from further analysis. Three of these finished the track before the memory task was completed. One participant was excluded due to technical issues with the simulator. Finally, one participant left the simulator to make a phone call. As a result, twenty-three participants (‘driving’: $n = 11$, ‘equal’: $n = 12$) were included in the present analysis. All of them responded to all ISA prompts in the experimental conditions. The driving task was performed as instructed, in that no obstacles were hit. The only exception was one driver who hit one obstacle, out of 186 obstacles passed. Demand transitions were first analyzed at the time frame of a full experimental condition, so that the diagnostic power of the NASA-TLX can be used to interpret ISA ratings. This was followed by an analysis at a trial block time frame. Manipulation checks of task difficulty and priority instructions were performed, using the same temporal distinction between full experimental conditions and trial blocks. This was due to an apparent absence of evident hysteresis. Finally, the impact of task prioritization is investigated.

5.3.1 Demand transitions between experimental conditions

The average duration of the transition from LOW_1 to $HIGH_2$ was 149.48 sec (*SE* = 7.42). This duration was measured from the end of the last ISA prompt in LOW_1 to the start of the first memory trial in $HIGH_2$. The second transition from $HIGH_2$ to LOW_3 (*M* = 106.26 sec, *SE* = 4.23) took significantly less time, $F(1,21) = 33.32$, $p < .001$, $\eta^2_p = .61$. This perhaps indicates a learning effect with respect to completing the NASA-TLX.

If demand transitions induced hysteresis, then performance and/or mental workload after $HIGH_2$ should differ from before $HIGH_2$. Panels A, C, E, and G in Figure 5.3 show that task performance and ISA ratings were similar in both LOW_3 and LOW_1, regardless of the instruction. Mixed 2 (Instruction) x 2 (Period) ANOVAs confirmed these observations, such that no significant effects were found. Although the average NASA-TLX ratings (Figure 5.4A) are similar in the LOW_1 and LOW_3 conditions, the 'driving' instruction appears to show differences on the sub-scales physical demand (Figure 5.4C), temporal demand (Figure 5.4D), and frustration (Figure 5.4G). Within the ‘equal’ instruction, effort (Figure 5.4F) appears lower in LOW_3. However, a mixed 2 (Instruction) x 2 (Period) x 6 (TLX scales) MANOVA yielded no significant effect. These findings imply that if demand transitions caused hysteresis, then the duration of such hysteresis must be shorter than the full experimental duration (i.e., 8 minutes).

5.3.2 Demand transitions between trial blocks

If no effect can be established across an extended interval of time, the next question is whether such an effect is potentially more transient in nature? To examine this, we use trial blocks (i.e., four memory trials plus an ISA prompt), which serve to offer a higher temporal resolution to identify hysteresis in terms of both performance and workload. The right panels in Figure 5.3 show task performance and ISA ratings per trial block. No NASA-TLX ratings were obtained at this resolution. Therefore, linear regressions were run to evaluate how ISA ratings related to the unweighted average of the six NASA-TLX sub-scales. ISA variance was significantly

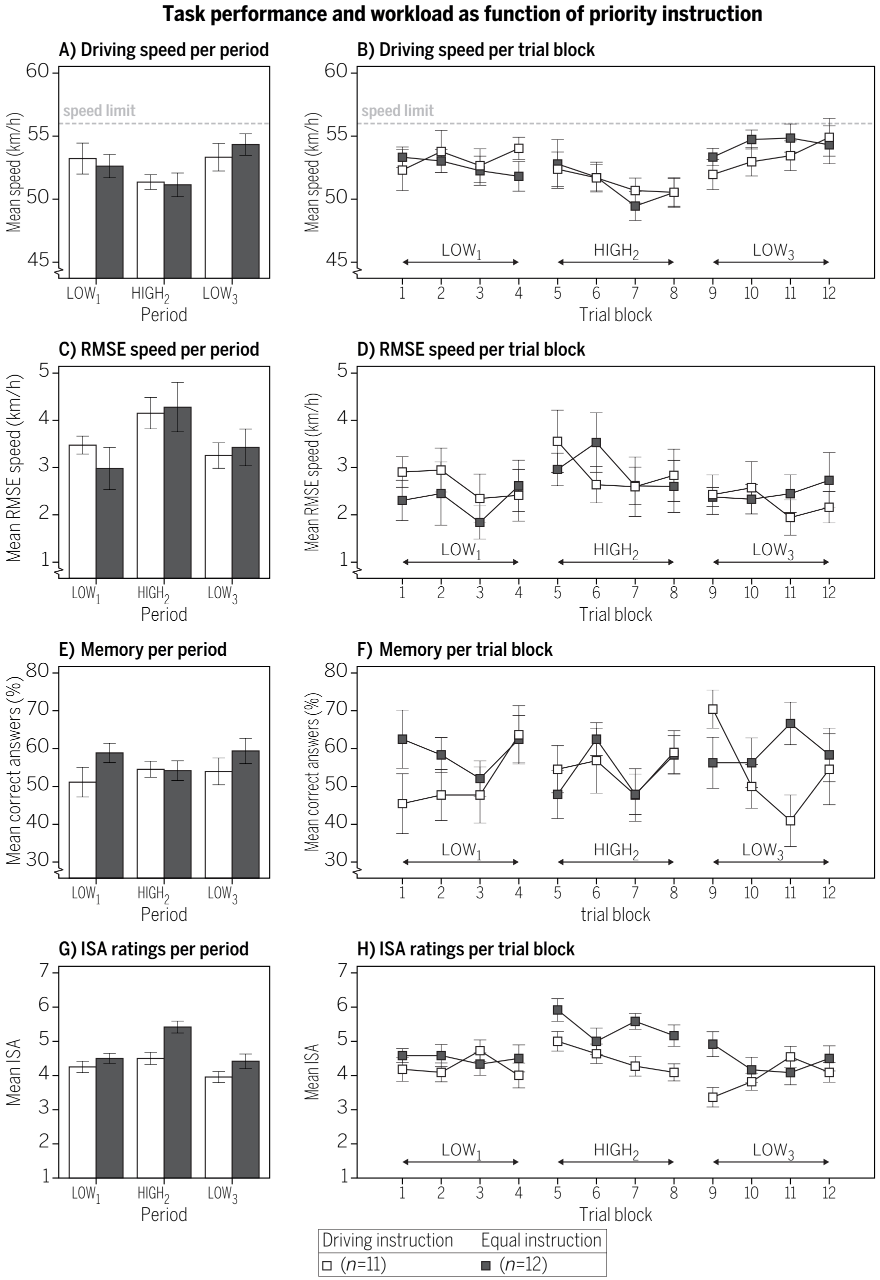

Figure 5.3. Task performance and workload as function of priority instruction. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability. NOTE: RMSE of driving speed is higher per period than per trial block, because it is calculated over a larger time frame.

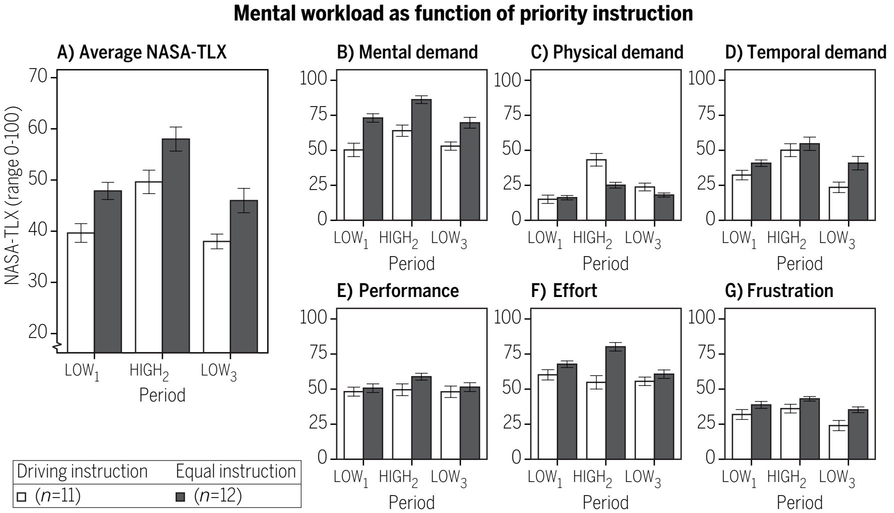

Figure 5.4. Unweighted average NASA-TLX and NASA-TLX sub-scales as function of priority instruction. Error bars represent +/- 1 standard error of the mean, corrected for within-subjects variability.

explained by TLX in LOW$_3$, $F(1,21) = 4.86$, $p < .05$, $R^2 = .19$, $\beta_{\text{std}} = .43$, $t(22) = 2.21$, $p < .05$, but not in the other experimental conditions. The latter suggests that ISA ratings should not be interpreted in terms of overall workload. Multiple linear regressions were run to investigate whether the sub-scales could explain the ISA ratings. The stepwise method excluded five sub-scales in each experimental condition. In HIGH$_2$ the variance of ISA was significantly explained by effort, $F(1,21) = 6.25$, $p < .05$, $R^2 = .23$, $\beta_{\text{std}} = .48$, $t(22) = 2.50$, $p < .05$. Furthermore, mental demand explained a significant amount of ISA variance in the LOW$_1$, $F(1,21) = 6.98$, $p < .05$, $R^2 = .25$, and in the LOW$_3$, $F(1,21) = 7.31$, $p < .05$, $R^2 = .26$, conditions. The analyses showed that mental demand significantly predicts ISA ratings in LOW$_1$, $\beta_{\text{std}} = .50$, $t(22) = 2.64$, $p < .05$, and in LOW$_3$, $\beta_{\text{std}} = .51$, $t(22) = 2.70$, $p < .05$. These findings indicate that ISA ratings during LOW$_1$ and LOW$_3$ can be interpreted in terms of mental demand.

The shortest time frame to identify hysteresis can be established by comparing the last trial block of LOW$_1$ (i.e., block 4) with the first trial block of LOW$_3$ (i.e., block 9). Participants with the 'driving' instruction showed decreasing driving speeds from block 4 to 9 (Figure 5.3B), as well as increasing memory performance (Figure 5.3F) with decreasing ISA ratings (Figure 5.3H). Conversely, participants with the 'equal' instruction showed increasing driving speed, decreasing memory performance, and increasing ISA ratings. However, neither of these interactions between Block and Instruction proved to be significant at this juncture. RMSE of driving speed (Figure 5.3D) appears stable from block 4 to 9, which was confirmed through statistical analysis. The only significant effect was found on ISA ratings. Participants with the 'equal' instruction reported higher mental workload than participants with the 'driving' instruction, $F(1,21) = 5.21$, $p < .05$, $\eta^2_p = .20$. However, this finding was not related to hysteresis per se. To summarize, no hysteretic effects were distinguished at a trial block time frame.

5.3.3 Manipulation check of experimental conditions

The absence of hysteresis raises the question whether the prerequisites to identify such effects were met. Most importantly, the manipulation of driving task difficulty should be reflected in task performance and/or mental workload. A 3 (Period) x 2 (Instruction) mixed ANOVA did not yield significant effects on task performance. However, a significant effect of Period on ISA ratings was found, $F(2,42) = 7.09$, $p < .01$, $\eta^2_p = .25$. ISA ratings increased significantly from LOW_1 to $HIGH_2$, $F(1,21) = 9.29$, $p < .01$, $\eta^2_p = .31$, and decreased from $HIGH_2$ to LOW_3, $F(1,21) = 10.77$, $p < .01$, $\eta^2_p = .34$ (see Figure 5.3G).

Figure 5.4A shows that the unweighted NASA-TLX average increased from LOW_1 to $HIGH_2$, and then subsequently decreased from $HIGH_2$ to LOW_3. Furthermore, the 'equal' instruction appears to induce greater mental workload than the 'driving' instruction. Panels B through G in Figure 5.4 suggest that the dual-task combination induced considerable mental demand and effort, but not so much physical demand or frustration. Multivariate results of a 3x2x6 MANOVA yielded a significant main effect of Period, Wilks' lambda = .37, $F(12,74) = 3.95$, $p < .001$, $\eta^2_p = .39$, and a significant interaction between Period and Instruction, Wilks' lambda = .55,

Table 5.1. Results of univariate ANOVAs on the NASA-TLX sub-scales.

NASA-TLX		**Univariate ANOVA**			**Contrast analysis (type: repeated)**				
subscale	**Source**	***df***	***F***	η^2_p	***df***	***F***	η^2_p	***F***	η^2_p
Mental demand	P	(2,42)	6.26 **	.23	(1,21)	8.56 **	.29	11.77 **	.36
Physical demand	P	(1.45,30.37)	16.89 ***	.45	(1,21)	23.52 ***	.53	12.80 **	.38
Temporal demand	P	(1.59,33.33)	9.18 **	.30	(1,21)	12.03 **	.36	11.06 **	.35
Frustration	P	(2,42)	4.30 *	.17	(1,21)	*n.s.*	-	10.19 **	.33
Physical demand	P × I	(1.45,30.37)	4.42 *	.17	(1,21)	6.35 *	.23	*n.s.*	-
Effort	P × I	(2,42)	3.61 *	.15	(1,21)	*n.s.*	-	5.13 *	.20

NOTE: P = Period, I = Instruction. The *df* of physical demand and temporal demand were adjusted using Greenhouse-Geisser, $\varepsilon = .72$ and $\varepsilon = .79$, respectively. Only significant effects are reported. * $p < .05$, ** $p < .01$, *** $p < .001$.

$F(12,74) = 2.18$, $p < .05$, $\eta^2_p = .26$. No significant main effect of Instruction was found, however.

Univariate ANOVAs showed that the main effect of Period was significant on four sub-scales (see Table 5.1), these being: mental demand, physical demand, temporal demand, and frustration. Contrast analyses revealed a significant increase from LOW$_1$ to HIGH$_2$ and a significant decrease from HIGH$_2$ to LOW$_3$ for mental demand, physical demand, and temporal demand. Frustration only increased significantly from LOW$_1$ to HIGH$_2$. The interaction between Period and Instruction was significant for physical demand and effort. Physical demand increased from LOW$_1$ to HIGH$_2$, but this increase proved to be larger in the 'driving' instruction (see Figure 5.4C). Effort decreased from HIGH$_2$ to LOW$_3$, but only in the 'equal' instruction group (see Figure 5.4F). To summarize, across the time frame represented by an experimental condition, the manipulations of driving task difficulty and priority instruction proved to be reflected in mental workload. However, no effects were found on task performance.

5.3.4 Manipulation check of trial blocks

While the previous manipulation check was appropriate for the time frame of an experimental condition, it is insufficiently discriminative for shorter durations. The right panels of Figure 5.3 show detailed patterns of task performance and ISA ratings at the resolution of a trial block. The manipulation of driving task difficulty should be visible in the first transition from LOW_1 to $HIGH_2$ (i.c., trial blocks 4 vs. 5) and the second transition from $HIGH_2$ to LOW_3 (i.c., trial blocks 8 vs. 9). However, 2 (Block) x 2 (Instruction) mixed ANOVAs on each transition did not yield significant effects for task performance. The only significant effect was found in the first transition of the ISA ratings (Figure 5.3H), which increased from trial block 4 to 5, $F(1,21) = 8.10$, $p < .05$, $\eta^2_p = .28$.

Surprisingly, no significant effects on ISA ratings were found in the transition from $HIGH_2$ back to LOW_3. Figure 5.3H suggests that ISA ratings decrease during $HIGH_2$, resulting in only a marginal difference between trial blocks 8 and 9. To analyze each experimental condition in isolation, separate 4 (Block) x 2 (Instruction) mixed ANOVAs were conducted for all measures on each of the three experimental conditions. A significant main effect of Block was found on ISA ratings during $HIGH_2$, Mauchly's test, $\chi^2(5) = 15.55$, adjusted using Huynh-Feldt, $\varepsilon = .75$, $F(2.25,47.31) = 3.13$, $p < .05$, $\eta^2_p = .13$. This effect was significant only from trial blocks 5 to 7, $F(1,21) = 5.05$, $p < .05$, $\eta^2_p = .19$, and from trial blocks 5 to 8, $F(1,21) = 14.33$, $p < .01$, $\eta^2_p = .41$.

Furthermore, Figure 5.3H shows that ISA ratings during LOW_3 increase with the 'driving' instruction, whereas they decrease with the 'equal' instruction. A significant interaction between Block and Instruction was found, Mauchly's test, $\chi^2(5) = 14.43$, adjusted using Huynh-Feldt, $\varepsilon = .80$, $F(2.41,50.62) = 3.59$, $p < .05$, $\eta^2_p = .15$. Simple contrasts revealed that this interaction was significant from trial blocks 9 to 11, $F(1,21) = 9.93$, $p < .01$, $\eta^2_p = .32$, and from trial blocks 9 to 12, $F(1,21) = 5.07$, $p < .05$, $\eta^2_p = .20$. Memory performance (Figure 5.3F) appears to follow an inverse trend of the ISA ratings during LOW_3: memory performance decreases with the 'driving' instruction, whereas it slightly increases with the 'equal instruction. This observation was supported by a significant interaction between Block and Instruction, $F(3,63) = 3.00$, $p < .05$, $\eta^2_p = .13$, which was significant from trial block 9 to 11, $F(1,21) = 10.12$, $p < .01$, $\eta^2_p = .33$. Lastly, trends in driving speed appear to decrease during $HIGH_2$ and increase during LOW_3, regardless of instruction, but a 4x2 ANOVA did not yield any further significant effects.

5.3.5 Impact of task prioritization

The manipulation of priority instructions was reflected in mental workload through an interaction effect with Period. Surprisingly this manipulation was not reflected in task performance across the time frame of an experimental condition. The interview results may explain why this was the case. One participant reported that following the 'driving' instruction they felt *"normal, like regular driving."* Another participant with an 'equal' instruction reported: *"It was easy to drive, but hard to listen. I concentrated more on the driving task."* These statements suggest that participants had preferences regarding task prioritization, which did not always accord directly with the instruction set they received.

To examine whether this conflict played a role, preferences were elicited and then used as an additional between-participant factor. Three raters independently evaluated the interview results to assign a post-hoc attribution of a 'driving' or 'equal' preference to each participant. The average agreement between the raters was 71%. Within the 'driving' instruction, 10 participants had a 'driving' preference, whereas 1 participant had an 'equal' preference. Within the 'equal' instruction, a 'driving' preference was found for 8 participants, whereas 4 participants had an 'equal' preference.

A 2 (Preference) x 2 (Instruction) x 3 (Period) mixed ANOVA was conducted on all measures. The participant with a 'driving' instruction and an 'equal' preference was excluded from this analysis, because there was no variance within this combination. The only significant effect was a main effect of Preference on memory performance, $F(1,19) = 5.17$, $p < .05$, $\eta^2_p = .21$. A 'driving' preference ($M = 53.47\%$, $SE = 2.87$) resulted in lower memory performance than an 'equal' preference ($M = 68.75\%$, $SE = 6.42$). An examination of hysteretic effects through three-way mixed ANOVAs did not yield significant effects involving Period (i.e., conditions LOW_1 vs. LOW_3), nor involving Block (i.e., trial blocks 4 vs. 9).

5.4 Discussion

The main results of this study are that an hysteretic effect in mental workload was found within the high demand condition, and contrary to what is commonly reported, no hysteretic effects were observed after the high-to-low demand transition. The latter observation is based on comparisons between relatively long experimental conditions (i.e., a time frame of 8 min), as well as between relatively short trial blocks (i.e., time frames of 2 min). The shortest hysteretic effect that could have been detected with the present setup corresponds with the transition time between the last two experimental conditions (i.c., 106 sec) plus the duration of a trial block (i.c., 120 sec). From prior work we can postulate that hysteresis likely occurs after a high-to-low demand transition. Thus, if hysteresis took place after the transition from $HIGH_2$ to LOW_3, it must have lasted less than 226 seconds.

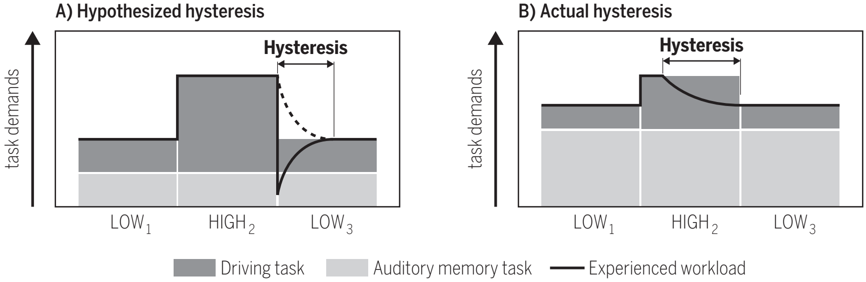

Figure 5.5. Hypothesized hysteresis (A) and actual hysteresis (B) resulting from demand transitions between experimental conditions. In panel A the solid line represents hypothesized hysteresis based on Hancock et al. (1995), and the dashed line is based on Morgan and Hancock (2011).

It has been assumed that the present study incorporates two demand transitions large enough to cause hysteresis after the second demand transition (see Figure 5.5A). However, the present findings challenge this assumption. The auditory memory task may have contributed more to the experienced workload than the driving task, resulting in a limited contrast between LOW_1 and LOW_3 on the one hand, and $HIGH_2$ on the other hand (see dark grey blocks in Figure 5.5B). Support for the contribution of the auditory memory task is found in the fact that the NASA-TLX mental demand sub-scale was rated higher than physical demand. The limited contrast follows from the observation that the manipulation of driving task difficulty has resulted in a difference that spans only 12% of the averaged NASA-TLX, and a corresponding degree of 11% of the ISA scale. Furthermore, participants may have habituated to the increased driving task demands during $HIGH_2$, especially because driving is an over-learned, everyday task. While the average ISA rating in $HIGH_2$ was significantly higher than the other experimental conditions, the ratings during $HIGH_2$ were actually decreasing. Consequently, ISA ratings returned to the level of LOW_1 at the beginning of LOW_3. This implies that hysteresis had already taken place immediately after the first demand transition (see solid line in Figure 5.5B).

It is unlikely that resource depletion caused the observed hysteresis in ISA ratings, because task demands were presumably constant throughout $HIGH_2$. The effort regulation theory, which has explained hysteresis in performance in previous work, may also explain hysteresis in mental workload. Our findings support this view. Driving performance and memory performance were unaffected by driving task demand across experimental conditions, whereas mental workload increased with driving task demand. The only NASA-TLX sub-scale unaffected by driving task demand was own performance, which apperception was consistent with actual performance. In addition, the effort NASA-TLX sub-scale proved to be the main predictor for ISA ratings in $HIGH_2$. Such findings are consistent with models featuring effort-related adjustment of attentional capacity (Hancock & Warm, 1989; Hockey, 1997). It appears participants improved the efficiency of their coping strategies during $HIGH_2$, to the extent that similar effort to LOW_1 and LOW_3 was required by the end of $HIGH_2$.

These findings have several theoretical and methodological implications for future research on demand transitions and resilience (and see Hoffman & Hancock, 2016). First, the observed ISA trends during experimental conditions demonstrate the importance of using on-line workload ratings in addition to data collected after the cessation of each experimental condition. Second, the inter-measurement time across demand transitions should be minimized, especially if hysteretic effects prove to have a short duration. Physiological measures may help identifying shifts in workload at an earlier instant than ISA ratings. Furthermore, the time it takes to complete a NASA-TLX questionnaire may be inappropriate in such circumstances, even though our results suggest that TLX completion time can itself be greatly reduced through practice. However, the NASA-TLX results have provided valuable information about the nature of experienced workload. The challenge, therefore, is to develop an on-line rating scale that strikes an appropriate balance between low obtrusiveness (i.e., the uni-dimensional ISA), and high diagnosticity (i.e., the multi-dimensional NASA-TLX).

An additional finding of this study is that the manipulation of priority instructions is reflected in mental workload, but not in task performance across the time frame of an experimental condition. As a result, the resource depletion and effort regulation theories could not be exhaustively tested as intended through task prioritization. An inquiry into prioritization preferences showed that participants with a preference for the driving task had lower memory performance than those with the 'equal' preference. Therefore, the instructions may have resulted in a significant effect on memory performance, if they matched with preferences. A dual-task study by Jansen et al. (2016) with a low-fidelity driving game suggests that such preferences are guided by judgments on task utility, and that conflicting instructions are followed only after extensive dual-task exposure. The present study shows that such a conflict may also play a role in the context of high-fidelity driving simulators. This warrants further research into the effect of preferences in combination with other secondary tasks.

Acknowledgments

We thank Jessica Milos, Emily Simpson, and Gabriel Yeager for their assistance in collecting the data, Tamar Nir for programming support, and Holly Robbins for recording the memory questions.

Key points

- We investigated carryover workload effects (hysteresis) in participants driving in a simulator. Participants drove a low workload roadway segment, followed by a high, and a low workload roadway stretch.
- The uni-dimensional Instantaneous Self Assessment (ISA) scale obtained subjective mental workload during experimental conditions, and the NASA-TLX was used afterwards. The relatively non-intrusive ISA proved useful for revealing the evolution of an hysteretic effect in mental workload, whereas the diagnostic power of the NASA-TLX served to interpret this effect.
- The temporal pattern of hysteresis in mental workload informs an appropriate timing of information presentation by communication systems.
- Analysis of verbal reactions suggested that pre-existing preferences regarding task prioritization conflicted with priority instructions. Such preferences may be useful to inform personnel recruitment and selection.

References

Anderson, G.S., Courtney, A., Plecas, D. & Chamberlin, C. (2005). Multi-tasking behaviors of general duty police officers. *Police Practice and Research*, 6(1), 39-48.

Borglund, E. & Nuldén, U. (2012). Personas in uniform: police officers as users of information technology. *Transactions on Human-Computer Interaction*, 4, 92-106.

Caird, J.K., Willness, C.R., Steel, P. & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4), 1282-1293.

Cannon, W.B. (1932). The wisdom of the body. New York: Norton.

Cox-Fuenzalida, L.-E. (2007). Effect of Workload History on Task Performance. *Human Factors*, 49(2), 277-291.

Cumming, R.W. & Croft, P.G. (1973). Human Information Processing under Varying Task Demand. *Ergonomics*, 16(5), 581-586.

Dressel, J. & Atchley, P. (2008). Cellular phone use while driving: A methodological checklist for investigating dual-task costs. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(5), 347-361.

Farrell, P.S.E. (1999). The hysteresis effect. *Human Factors*, 41(2), 226–240.

Girard J.M., Wilczyk M., Barloy Y., Simon P. & Popieul J.C. (2005). Towards an on-line assessment of subjective driver workload. *Driving Simulation Conference North America* (pp. 382-391), Orlando, FL.

Gluckman, J.P., Warm, J.S., Dember, W.N. & Rosa, R.R. (1993). Demand Transitions and Sustained Attention. *The Journal of General Psychology*, 120(3), 323-337.

Goldberg, R.A. & Stewart, M.R. (1980). Memory overload or expectancy effect? Hysteresis revisited. *Ergonomics*, 23(12), 1173-1178.

Hancock, P.A., Williams, G. & Manning, C.M. (1995). Influence of Task Demand Characteristics on Workload and Performance. *The International Journal of Aviation Psychology*, 5(1), 63-86.

Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31(5), 519-537.

Hart, S.G. & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-184). Amsterdam, The Netherlands: North Holland Press.

Hill, S.G., Iavecchia, H.P., Byers, J.C., Bittner, A.C., Zaklade, A.L. & Christ, R.E. (1992). Comparison of Four Subjective Workload Rating Scales. *Human Factors*, 34(4), 429-439.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1-3), 73-93.

Hoffman, R.R. & Hancock, P.A. (2016). Measuring resilience. *Manuscript submitted*.

Jansen, R.J., Van Egmond, R., De Ridder, H. & Silvester, S. (2014). Transitional Journey Maps: Capturing the dynamics of operational policing. In: De Waard, D., Brookhuis, K., Wiczorek, R., Di Nocera, F., Barham, P., Weikert, C., Kluge, A., Gerbino, W. & Toffetti, A. (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference* (pp. 15-27). Torino, Italy.

Jansen, R.J., Van Egmond, R., De Ridder, H. (2016). Task prioritization in dual-tasking: instructions versus preferences. *Manuscript submitted*.

Jordan, C.S. (1992). Experimental study of the effect of an instantaneous self assessment workload recorder on task performance. *DRA Technical Memorandum (CAD5) 92011*. DRA Maritime Command Control Division, Portsmouth.

Kantowitz, B. H. & Knight, J. L. (1976). Testing tapping timesharing. II: Auditory secondary tasks. *Acta Psychologies*, 40, 343-362.

Leggatt, A. (2005). Validation of the ISA (Instantaneous Self Assessment) subjective workload tool. In: Bust, P.D. & McCabe, P.T. (Eds.), *Proceedings of the International Conference on Contemporary Ergonomics* (pp. 74-78), 5-7 April 2005, Hatfield, UK.

Lemercier, C., Pêcher, C., Berthié, G., Valéry, B., Vidal, V., Paubel, P.-V., Cour, M., Fort, A., Galéra, C., Gabaude, C., Lagarde, E. & Maury, B. (2014). Inattention behind the wheel: How factual internal thoughts impact attentional control while driving. *Safety Science*, 62, 279-285.

Matthews, M.L. (1986). The influence of visual workload history on visual performance. *Human Factors*, 28(6), 623-632.

Matthews, M.L. & Desmond, P. (2002). Task-induced fatigue states and simulated driving performance. *Quarterly journal of experimental psychology*, 55(2), 659-686.

Morgan, J.F. & Hancock, P.A. (2011). The effect of prior task loading on mental workload: An example of hysteresis in driving. *Human Factors*, 53(1), 75-86.

Norman, D.A. & Bobrow, D.G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44-64.

Reid, G.B. & Nygren, T.E. (1988). The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload. In: P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 185–218). Amsterdam, The Netherlands: Elsevier Science.

Sawyer, B.D., & Hancock, P.A. (2012). Development of a linked simulatino network to evaluate intelligent transportation system vehicle to vehicle solutions. In: *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (pp. 2316-2320). Santa Monica, CA.

Sørensen, C. & Pica, D. (2005). Tales from the police: rhythms of interaction with mobile technologies. *Information and Organization*, 15, 125-149.

Tattersall, A.J. & Foord, P.S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, 39(5), 740-748.

Ungar, N.R., Matthews, G., Warm, J.S., Dember, W.N., Thomas, J.K., Finomore, V.S. & Shaw, T.H. (2005). Demand Transitions and Tracking Performance Efficiency: Structural and Strategic Models. In: *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting* (pp. 1523-1526). Santa Monica, CA.

Zar, J.H. (1996). *Biostatistical Analysis*. 3rd ed. Upper Saddle River, New Jersey: Prentice Hall.

Chapter 6

Conclusion

Task prioritization is the process of allocating attention to one task (e.g., listening to the radio) at the expense of another task (e.g., driving a car) (Gopher et al., 1989). Task prioritization is an important skill in operational policing, a dynamic multi-task context in which police officers are known to suffer from work overload. Good design can help reduce work overload by connecting with peoples' skills and abilities (Norman, 2010). Therefore, designers of information technologies benefit from knowledge on task prioritization.

The scientific goal of this thesis has been to understand the mechanisms that underlie and/or result from task prioritization in complex socio-technical systems, such as the police context. An applied goal has been to investigate the leeway to push information to police officers while driving their vehicle in varying work situations. To pursue these goals, an observation study on the dynamics of operational policing has initiated five dual-task experiments with a driving task and an auditory memory task.

Several contributions have resulted from this endeavor. A novel method has been presented to describe the dynamics of complex socio-technical systems. Theoretical contributions include the discovery of preferences in task prioritization, and elaborations on two existing theories on dual-task execution. Furthermore, methodological recommendations on online measurement of mental effort have been presented.

This final chapter presents an overview of the key findings in Section 6.1. Practical implications for the police context are provided in Section 6.2. Finally, limitations and recommendations for further research are discussed in Section 6.3.

6.1 Key findings

Four mechanisms related to task prioritization, task demand, task performance, and mental effort have been described across the chapters of this thesis: workflow fragmentation, preferences regarding task prioritization, coping strategies, and hysteresis.

6.1.1 Workflow fragmentation

At a macro temporal resolution (i.e., minutes to hours), Chapter 2 has introduced the Transitional Journey Map (TJM), a method to visualize social and socio-technical interactions in an unpredictable workflow.

> **Key finding 1:** Transitional Journey Maps are an effective method to reveal workflow fragmentation.

Dutch operational police work has been captured in terms of six activity categories (i.c., 'engage at incident', 'driving to incident', 'driving surveillance', 'driving to station', 'parked surveillance', and 'parked at station'). Transitions between these activity categories occur frequently. Consequently, a visualization of the transitions provides an immediate understanding of the degree to which the workflow is fragmented. Additionally, workflow fragmentation has been quantified by describing the duration of segments of activities. The resulting median duration within each activity category has proven to be much shorter than the overall time spent on an activity, as is typically reported in tables (e.g., Anderson et al., 2005; Frank et al., 1997; Smith et al., 2001). This demonstrates that a high level aggregation of activity durations does not give a proper representation of the dynamics of operational policing. Furthermore, it raises the question what causes transitions.

> **Key finding 2:** Qualitative data is essential for an understanding of quantitative data on workflow fragmentation.

In the context of operational policing, transitions toward the 'driving to an incident' activity category are typically accompanied by incoming messages from the dispatcher. By including verbal data, TJMs have enabled an examination of the cause for transitions between activity categories. Furthermore, retrospective accounts allow for triangulation with observed events that lack momentary verbal data. In addition to the police context, the TJMs have also been used to unravel alarm response behavior in satellite control rooms. Without verbal data, it would not have been possible to understand why some alarms do not elicit a response by the control room operator.

> **Key finding 3:** Dutch police officers want to receive more information than they can process.

Qualitative data analysis has confirmed that monitoring incoming radio messages is an essential task in operational policing, which has to be executed even when driving. Some police officers even wish to monitor three police radio channels simultaneously. Therefore, police work requires that task prioritization differs from regular drivers. At the same time, police officers frequently have to ask the dispatcher to repeat a message, because they struggle to remember its contents. Comments on this difficulty have been accompanied by reports on work overload in all activity categories, except 'driving to station'. To summarize, the information needs and processing abilities by Dutch police officers are not in balance. This warrants a redesign of the information technologies in police vehicles.

6.1.2 Preferences

At a micro temporal resolution (i.e., milliseconds to seconds), Chapter 3 has addressed the observation that police officers differ in the way they assign priority levels to the driving task and the monitoring task, even though the occupational requirements for task prioritization (i.e., on a macro level) are identical. These differences have been interpreted in terms of preferences regarding task prioritization. In an experimental setting some people prefer to perform the driving task (i.e., a 'driving' preference), whereas other people prefer to perform both the driving task and the auditory task (i.e., an 'equal' preference). The distribution of these preferences is approximately 2:1 in favor of the 'driving' preference. Preferences appear to have been guided by judgments on the utility of each task, and they are reflected in task performance. Are priority instructions (i.e., occupational requirements in operational policing) always followed, considering that there may be conflicting preferences?

> **Key finding 4:** People have preferences regarding task prioritization in a dual-task situation. The impact of preferences on whether priority instructions are followed diminishes with increasing dual-task exposure.

This finding has been explained by integrating two theories: Threaded Cognition Theory (TCT) (Salvucci & Taatgen, 2008), and the Compensatory Control Model (CCM) (Hockey, 1997, 2011), see Figure 6.1. TCT explains dual-tasking in terms of rapid switching between task goals. However, TCT does not address how priority levels are set. This is where the CCM comes into play, which explains the prioritization of a task goal as function of a utility judgment. If the execution of a task goal is jeopardized by changes in external task demands (i.e., 'demand transitions' in Figure 6.1, and see Chapters 4 and 5), and the task goal is regarded as important (i.e., due to 'preferences' or 'instructions' in Figure 6.1), then one has two options to protect performance. The first option is to increase mental effort. The second option is to increase the priority level of that task goal relative to the priority levels of other task goals. It should be noted, though, that CCM

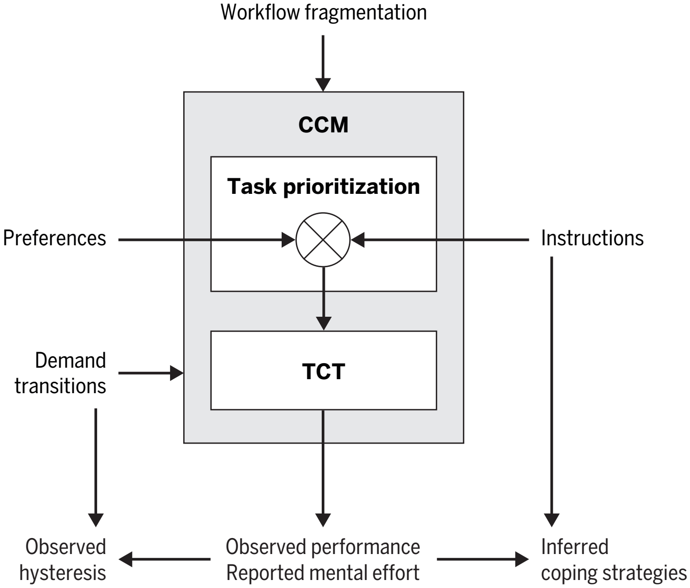

Figure 6.1. Schematic overview of research concepts in this thesis (adapted from Figure 3.10). Workflow fragmentation influences which tasks are applicable to a given situation. The Compensatory Control Model (CCM) describes how tasks are prioritized on a seconds to minutes temporal resolution. Threaded Cognition Theory (TCT) describes rapid switching between task goals on a millisecond level, as function of the priority levels set by TCT.

does not address how dual-tasking takes place, contrary to TCT. Therefore, an integration of TCT and CCM in Chapter 3 has combined the explanatory powers of each theory, and has mitigated their short-comings.

6.1.3 Coping strategies

At a meso temporal resolution (i.e., seconds to minutes), Chapter 4 has addressed reports of work overload by investigating whether differences in task prioritization result in distinct coping behavior. This investigation has been guided the observation that police officers on patrol are expected to pay more attention to non-driving tasks (e.g., auditory monitoring) than regular drivers, potentially at the expense of attending the driving task. Hockey’s (1997, 2011) CCM predicts five coping strategies when task demands increase, where coping strategies are formulated in terms of a tradeoff between task performance and mental effort. Chapter 4 has postulated the logics to infer these coping strategies from task prioritization and the above tradeoff (i.e., 'Inferred coping strategies' in Figure 6.1). These logics have been implemented in an experiment.

> **Key finding 5:** Task prioritization influences coping behavior in the dual-task context of driving with an auditory task.

People have no problem in applying a coping strategy when the driving task has to be prioritized. However, when both tasks have to be prioritized equally, people fail to apply a coping strategy: task performance decreases on both tasks, while mental effort increases. These findings suggest that task prioritization plays a role in work overload at operational policing.

> **Key finding 6:** Two new coping strategies have been discovered, namely 'intense focus' and 'exclusive decrement'.

The postulated logics to infer coping strategies predict the existence of two new coping strategies that are not currently defined by the CCM. First, the 'intense focus' strategy corresponds with the situation in which primary task demand is so high, that despite neglecting the secondary task, additional effort is still required to protect primary task performance. Second, the 'exclusive decrement' strategy concerns the protection of mental effort and performance on one task, at the expense of another task. The 'intense focus' strategy is only applicable when one task is clearly prioritized over another task. On the other hand, the 'exclusive decrement' strategy is only applicable when two tasks are equally prioritized (i.e., one can no longer speak of 'primary' and 'secondary' tasks). The existence of both coping strategies has been confirmed by the experiment described in Chapter 4.

6.1.4 Hysteresis

Chapter 5 also concerns the meso temporal resolution. Workflow dynamics in operational policing have been addressed by investigating hysteretic (i.e., carry-over) effects from sudden transitions in driving task demands (i.e., 'Demand transitions' in Figure 6.1). Two hypotheses have been tested;

one on the occurrence of hysteresis, and one on the effect of task prioritization on hysteresis.

The first hypothesis has predicted that a sudden transition from high-to-low task demands would result in lagged performance and/or mental effort, compared to a baseline condition with low task demands. Contrary to what had been expected, no hysteretic effects have been found following the high-to-low demand transition. However, an hysteretic effect in mental effort has been found following the initial low-to-high demand transition (i.e., 'Observed hysteresis' in Figure 6.1). A sudden increase in driving task demands was initially mirrored by an increase in mental effort, but then mental effort gradually dropped back to the baseline level before the high task demand condition ended. Therefore:

> **Key finding 7:** To describe hysteretic patterns in mental effort, repeated measurements of subjective mental effort should be obtained during experimental conditions, and not only afterwards.

The second hypothesis has predicted that task prioritization would influence in which task the above performance lag would occur, and how long it would take to normalize to the low task demand baseline level (i.e., 'hysteresis'). However, the second hypothesis could not be tested, because the priority instructions did not yield any differences across the experimental conditions. An explanation has been found from a post-hoc analysis, which has revealed conflicting preferences regarding task prioritization. In line with the experiments in Chapter 3, an 'equal' preference was accompanied by improved auditory memory performance compared to a 'driving' preference. This means that conflicting preferences have not only been found in a low fidelity driving game (i.e., Chapter 3), but also in a high fidelity driving simulator (i.e., Chapter 5). Therefore:

> **Key finding 8:** Experimental findings support the idea that preferences affect the extent to which priority instructions are followed in real-life driving situations.

6.2 Practical implications

This section addresses the applied goal of this thesis by focusing on practical implications for the police context. The starting point is the paradox that police officers want to receive more information than they can process (i.e., key finding 3), which in turn contributes to work overload. Based on the findings of this thesis, three organizational and technical approaches to resolve the paradox are evaluated: task distribution, filtering information, and timing of information.

6.2.1 Task distribution is not enough

The experiments reported in this thesis have focused on solo patrol, by involving only one participant at a time. In this setting, driving with an auditory memory task has required high levels of mental workload, especially when one has been instructed to prioritize both tasks equally (see Chapter 3 and 4). The cognitive resources of the socio-technical system may be increased by shifting from solo patrol to duo patrol (Stahl, 2011). Hampton & Langham (2005) argue that duo patrol increases productivity, whereas solo patrol is less effective and potentially compromises safe driving. Therefore, distributing the tasks across two police officers (i.e., a driver and a co-driver) appears to be a feasible solution at first sight. The driver would be expected to fully prioritize the driving task and to ignore incoming radio messages, which are now the responsibility of the co-driver. With the Dutch police, this is in fact the formal distribution of duo patrol work.

However, even in duo patrol situations, police officers with a driver role have reported work overload due to excessive radio communication (see Chapter 2). This suggests that officers are not able or willing to follow the formal priority instruction that comes with the driver role (cf. Chapter 3). Furthermore, Chapter 4 has shown that degradation of the signal-to-noise ratio of incoming messages results in increased mental effort, even when those messages should have been ignored in the first place (i.e., following a 'driving' priority instruction). In line with Levy & Pashler (2008), this questions to what extent police officers will be able to fully concentrate on the driving task. Therefore, task distribution across team members may alleviate the observed paradox, but it is not sufficient to fully mitigate its effect on work overload.

6.2.2 Filtering information: a word of caution

An alternative approach to counteract work overload is to reduce the amount of information that is to be processed. Chapter 2 has shown that the vast majority of incoming radio messages is not addressed to the recipient. Filtering those messages would reduce the demands of the monitoring task.

However, one should be cautious about unwanted side effects that may result from filtering information. An example of such side effects is reported by Meehan (1998). In his observation study on a U.S. police unit, he has studied the effect of selective communication (i.e., addressing one recipient) versus radio messages broadcasted to the entire unit. A negative side effect of selective communication was that officers not assigned to a call had little or no idea what was happening on the street. Consequently, it was difficult to estimate when colleagues required assistance. Similar findings have been reported in the observation study of Chapter 2. In addition the above functional role, these messages also offer a means to provide emotional support to colleagues. For example, when one of the observed police officers returned to the police station after a traumatic experience, his colleagues immediately asked how he was doing. As it turned out, they had all been listening in on their radios. This demonstrates that broadcasted (i.e., non-selective) messages play a fundamental role in the social structure of the police organization.

Chan (2001) argues that new communication technologies inevitably change police practice. In line with Chan, the above examples advocate that one should not redesign the technical part of a socio-technical system without addressing the impact on its social counterpart.

6.2.3 Timing of information

A survey by Colvin and Goh (2005) shows that timely access to information and information quality (e.g., relevance, accuracy, recency, specificity) are the most important components of technology acceptance by U.S. patrol officers. Furthermore, an interview by Mudde (2013) notes that Dutch police officers have often forgotten relevant information at an incident site, when they have had to remember this information for a prolonged period of time. Branaghan et al. (2010) have acknowledged this need for the right information at the right time by developing distinct graphical user interfaces for the Mobile Data Terminal. In their mock-up, the type of call (e.g., pursuit, theft, traffic stop) determines which information is presented, and in what layout. What has not been addressed, though, is in which circumstances it is feasible for police officers to process the information, where feasibility is expressed in terms of task performance and mental effort.

New insights can be derived from a comparison between Chapters 3 and 4. Experiment 1 in Chapter 3 has showed that, initially, less destinations are reached when adding the auditory memory task. This has resulted in an average time delay of 13.8%[4]. Such a delay is unacceptable for emergency response driving, where a swift response is crucial. However, no such delay has been found in the driving task in Chapter 4, regardless of the type of route, nor the presence of distortion in the auditory stimuli. Therefore, this difference may be explained by the characteristics of the driving tasks in both experiments. In Experiment 1 of Chapter 3, participants continuously had to reach new destinations, which required them to plan their own route. This navigation component was absent in Chapter 4, where

---

[4] This delay has been calculated as follows. Participants in the Early DUAL Sequence performed equal during the $DR_{baseline}$ (M = 58.44%) and DUAL (M = 58.44%), the latter of which included the auditory memory task. However, performance in the Late DUAL Sequence increased from $DR_{baseline}$ (M = 62.86%) to DRrepeat (M = 72.86%), without an additional task. This increase corresponds with 1.4 destinations. With a grand mean destination duration of 29.6 seconds, this yields a delay of 41.4 seconds, or 13.8% of the total task duration.

participants always drove a fixed route. In other words, participants may have been more familiar with their driving task environment in Chapter 4. They may have executed the driving task as a routine task, which enabled them to direct more mental effort towards the auditory memory task, without degrading their driving performance.

The characteristics of the experimental driving tasks can be related to distinct sections of a police emergency response. Police officers initially use familiar main roads when driving from the police station to the neighborhood where an incident was reported (e.g., the familiar driving task in Chapter 4). Area knowledge of the neighborhood itself may be limited, however (i.e., the driving task in Chapter 3, and see Chapter 2). This requires police officers to direct effort towards finding the exact incident location. This means that for the shortest time-to-arrival, information on the incident is best provided as early in the drive as possible. A disclaimer is at place, though. If the aim is to minimize mental effort, and the driving task demands have just increased (e.g., traffic density), then the findings of Chapter 5 suggest that at least four minutes of habituation are required before the information can be pushed.

6.3 Limitations and recommendations

Limitations of the present research, as well as recommendations for further research, are clustered in three themes: the relation between the naturalistic and experimental settings, the integration of TCT within CCM, and the potential influence of workflow fragmentation on work overload.

6.3.1 Relation between naturalistic and experimental settings

This thesis is a marriage of two research methodologies: naturalistic observations and controlled laboratory experiments. An advantage of this approach is that the former methodology (i.e., Chapter 2) ensures that the experiments (i.e., Chapters 3 to 5) are grounded in real-life problems, thereby increasing their ecological validity. A disadvantage is that the time spent on one methodology is subtracted from the time available for the other methodology. Consequently, choices have been made with regard to how close the experimental settings could resemble their naturalistic counterpart.

A focus on dual-tasking, as opposed to multi-tasking, is one of those choices. It is known that police officers are frequently involved in multiple in-vehicle tasks while driving (Anderson et al., 2005; and see Chapter 2). The instruction to prioritize all tasks equally probably would have resulted in even higher mental effort ratings, had more tasks been used in Chapters 3 to 5. The choice to focus on dual-tasking has been made to limit the methodological complexity of the experiments. For future research on multi-tasking, the proposed goal selection mechanism in Chapter 3 should be extended. Furthermore, the presence of multiple in-vehicle tasks provides an opportunity for alternative task prioritization settings, both in terms of preferences and in terms of priority instructions. These alternatives should be incorporated in the rules with which coping strategies have been inferred in Chapter 4.

Another choice concerns the operationalization of the driving task. The closer the experimental driving task resembles police driving, the better one can project the experimental results on the police context. However, a high speed driving experiment with a real car in actual traffic is not feasible due to ethical, legal, and technical constraints. For the same reason, no attempt has been made to simulate the emotional stress that often accompanies emergency responses (e.g., when rushing towards a report of child abuse). With regard to the academic goal of this thesis, such a realistic setting is fortunately not required to examine preferences, coping strategies, and hysteresis. Therefore, a low-fidelity driving game has been used in Chapters 3 and 4, and a driving simulator has been used in Chapter 5 (note: the simulator featured the cockpit of an actual police car).

Is it possible, then, to generalize the driving task results to the police context? Santos et al. (2005) compared their results from laboratory, simulator, and (instrumented) field studies. The use of an in-vehicle information system was reflected in decreased driving performance in all settings. This finding suggests that the impact of the auditory memory task on driving performance in Chapters 3 to 5 can be generalized to the police context. Nevertheless, a word of caution is in order. Common measures on driving safety, such as lateral position, headway distance, and speed choice in relation to the speed limit, may be irrelevant to emergency response driving by police officers. In contrast, the primary goal of the driving task in Chapters 3 and 4 was to arrive as fast as possible at a designated area. Therefore, further research is required to verify the generalizability of the results of those chapters, preferably in a field setting.

Finally, choices have been made in the operationalization of the radio monitoring task. Spence and Ho (2008) argue that researchers should use auditory information processing tasks that are representative of what people hear in their daily lives. Such tasks have been found to place higher cognitive demands than simplistic and artificial auditory tasks, resulting in greater impairment on driver performance. In line with Spence and Ho, and aiming at generalization to the police context, police radio messages would have been the ideal stimuli for the memory task. However, students have been recruited for the experiments, because it was not feasible to recruit police officers. Students are not familiar with the protocols in police radio communication. News items, on the other hand, are part of their daily lives, and they have been used successfully in other studies (e.g., Ünal et al., 2013). The main limitation of using news items, is that they are not related to the driving context, whereas police radio messages often instruct officers towards a destination. Hence, failure to attend incoming messages has no impact on the navigation sub-task of the driving task. This limitation may have facilitated a bias in preference towards the driving task (i.e., 2:1 versus the 'equal' preference).

6.3.2 Integration TCT within CCM

The experiments of Chapter 3 that resulted in the proposed integration of TCT within CCM have three methodological limitations. First, the experiments have used a modest sample (i.e., *N* between 21 and 43 subjects). Second, the consequence of asking people afterwards about their preference, is that this procedure has caused unequal sample sizes, and low numbers in certain conditions. Third, and related to the above procedure, differences in task skills may have biased the reported preferences. A follow-up study should address these concerns, by using a larger sample size, a method to unobtrusively infer preferences a priori, and by setting personalized task demand levels (e.g., Somberg & Salthouse, 1982).

A theoretical limitation of the proposed integration is that it is not detailed enough to be implemented in the cognitive architecture in which TCT has been modeled. Consequently, validation is not yet possible. Several challenges lie ahead. One of them is to develop a connection from the fast switching mechanism of TCT to the relatively slow decision structure of CCM. Specifically, a model that predicts mental effort in dual-tasking is needed. Conversely, another challenge is to develop a connection from CCM to TCT. Here, the key issue is to understand how TCT uses the task priority levels set by CCM. Chapter 3 has provided a preliminary model that describes goal selection in a dual-task situation, but this model has not yet been tested. Finally, the decision structure of CCM has to be modeled. Inspiration may be drawn from the rule-based utility judgments described by Kurzban et al. (2013). The logics to infer coping strategies in Chapter 4 provide a means to verify such a model.

6.3.3 Workflow fragmentation and work overload

One of the goal of this thesis has been to understand how task prioritization mechanisms are related to the dynamics of a complex socio-technical system. Chapter 2 has described workflow fragmentation in terms of transitions between activity categories. Next, Chapters 3 to 5 have described the effects of (varying) task demands on performance and mental effort.

However, a link is still missing between performance, mental effort and transitions between activity categories, because sudden transitions in task demands are not necessarily related to transitions between activity categories. For example, during an emergency response officers may face fluctuations in traffic density. These fluctuations cause demand transitions in the driving task, but the whole drive takes place in only one activity category (i.e., ‘driving to incident’, no transition).

This link should be addressed by examining performance and mental effort both at a task level, and at an activity category level (cf. González & Mark, 2005). A challenge will be to distinguish between immediate and hysteretic effects resulting from a transition between activity categories. If the transition to another activity category structurally takes place before hysteretic effects of a former transition have ended, then it may appear as if the transitions have no effect on performance and mental effort. However, what may actually be the case is that workflow fragmentation has caused a situation of permanent work overload. The reports on work overload by police officers are an indication of such a situation. Therefore, further research on workflow fragmentation is needed to understand, and potentially reduce work overload.

References

Anderson, G.S., Courtney, A., Plecas, D. & Chamberlin, C. (2005). Multi-tasking behaviors of general duty police officers. *Police Practice and Research 6*(1): 39-48.

Branaghan, R.J., Takamura, J., Palmer, M.T., Hildebrand, E.A. & Sevier, D.C. (2010). Mapping Information Requirements for Police Patrol and Response to Informational Displays. *Journal of Cognitive Engineering and Decision Making 4*(2): 113-128.

Chan, J.B.L. (2001). The Technological Game:: How Information Technology is Transforming Police Practice. *Criminology and Criminal Justice 1*(2): 139-159.

Colvin, C.A. & Goh, A. (2005). Validation of the technology acceptance model for police. *Journal of Criminal Justice 33*(1): 89-95.

Frank, J., Brandl, S.G. & Watkins, R.C. (1997). The content of community policing: a comparison of the daily activities of community and 'beat' officers. *Policing: An International Journal of Police Strategies & Management 20*(4): 716-728.

Gopher, D., Weil, M. & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica 71*(1-3): 147-177.

Hampton, P. & Langham, M. (2005). A contextual study of police car telematics: the future of in-car information systems. *Ergonomics 48*(2): 109-118.

Hockey, G.R.J. (2011). A motivational control theory of cognitive fatigue. In: Ackerman, P.L. (Ed.) *Cognitive fatigue: multidisciplinary perspectives on current research and future applications*. Washington, DC, American Psychological Association: 167-188.

Hockey, G.R.J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology 45*(1-3): 73-93.

Kurzban, R., Duckworth, A., Kable, J.W. & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences 36*(6): 661-679.

Levy, J. & Pashler, H. (2008). Task prioritisation in multitasking during driving: opportunity to abort a concurrent task does not insulate braking responses from dual-task slowing. *Applied Cognitive Psychology 22*: 507-525.

Meehan, A. (1998). The Impact of Mobile Data Terminal (MDT) Information Technology on Communication and Recordkeeping in Patrol Work. *Qualitative Sociology 21*(3): 225-254.

Mudde, T. (2013). Agenten onthouden bar weinig van briefing, blijkt uit onderzoek [Research shows that police officers remember little from their briefing]. *Volkskrant*, 3 July 2013, http://www.volkskrant.nl/binnenland/agenten-onthouden-bar-weinig-van-briefing-blijkt-uit-onderzoek~a3469461/

Norman, D.A. (2010). *Living with complexity*. Cambridge, MA, MIT Press.

Salvucci, D.D. & Taatgen, N.A. (2008). Threaded cognition: an integrated theory of concurrent multitasking. *Psychological Review 115*(1): 101-130.

Santos, J., Merat, N., Mouta, S., Brookhuis, K. & de Waard, D. (2005). The interaction between driving and in-vehicle information systems: Comparison of results from laboratory, simulator and real-world studies. *Transportation Research Part F: Traffic Psychology and Behaviour 8*(2): 135-146.

Smith, B.W., Novak, K.J. & Frank, J. (2001). Community Policing and the Work Routines of Street-Level Officers. *Criminal Justice Review 26*(1): 17-37.

Somberg, B.L. & Salthouse, T.A. (1982). Divided attention abilities in young and old adults. *Journal of Experimental Psychology: Human Perception and Performance 8*(5): 651-663.

Spence, C. & Ho, C. (2008). Multisensory interface design for drivers: past, present and future. *Ergonomics 51*(1): 65-70.

Stahl, G. (2011). Theories of cognition in CSCW. In: Bødker, S., Bouvin, N.O., Wulf, V., Ciolfi, L. & Lutters, W. (Eds.) *ECSCW 2011: Proceedings of the 12th European Conference on Computer Supported Cooperative Work,* 24-28 September 2011, Aarhus Denmark. London, Springer London: 193-212.

Ünal, A.B., Platteel, S., Steg, L. & Epstude, K. (2013). Blocking-out auditory distracters while driving: A cognitive strategy to reduce task-demands on the road. *Accident Analysis & Prevention 50*: 934-942.

Summary

This dissertation presents the results of a series of studies on monitoring radio messages while driving, an omnipresent dual-task combination in police work, but also one that is considered unsafe for regular drivers. The focus is on task prioritization, the process of allocating attention to one task at the expense of another task (Gopher et al., 1989). Contrary to regular drivers, police officers typically do not have the option to stop their car to attend important incoming messages, nor can they afford an uninformed arrival at the scene. In other words, regular drivers are expected to fully prioritize the driving task, whereas police officers are expected to prioritize the driving and radio communication tasks with approximately equal priority. This expectation may explain recent reports on work overload. Therefore, the main goal of this dissertation is to understand the mechanisms that underlie and/or result from task prioritization in a dynamic complex socio-technical system, such as the police context. Generated insights from human cognition may aid the development of information technology for future police vehicles.

Chapter 2 explores how situational demands in police work influence the necessity of task prioritization. Operational police work can be characterized by the continuous switching between surveillance, responding to incidents, and office activities. Transitions between these

activities are initiated by radio contact, messages on a mobile data terminal, or personal observations. The "information environment" emerging from these channels may cause cognitive overload during demanding activities. Although the notion of fragmented work is acknowledged in police literature, detailed descriptions are lacking. The goal of this study is to better understand cognitive load in police officers by capturing the dynamics of operational policing. Ten officers of the Dutch police force were accompanied while on patrol with their car. The method of contextual inquiry was used to collect 28 hours of data. Activities were mapped on a pre-defined set of categories. Attention was paid to how officers experienced their information environment while performing these activities. All was captured in the Transitional Journey Map, a new method to visualize workflow. The Transitional Journey Map augments a sequence of activities with experiential and contextual information. This method was used to identify cognitive overload situations and differences between solo and dual patrol work. These insights are relevant for improving the information system that assists officers in their patrol vehicle.

In addition, Chapter 2 describes an observation study in satellite control rooms to examine whether the Transitional Journey Map method can be applied across different domains. Five operators in three satellite control rooms were observed to assess the functionality of their alarm design. A total of 31 alarm signals were recorded, of which only two resulted in problem-solving behavior. Transitional Journey Maps were used to represent qualitative interview data and observation data. Statement analysis revealed that operators anticipated the majority of the alarms. Implications for alarm signal design are discussed. Furthermore, a comparison of these observation studies shows that the process of workflow visualization shows similarities with design sketching. An important take-away message is that a priori categorization is not essential, as long as data is collected on multiple levels of abstraction.

The role of task prioritization in performance tradeoffs during multitasking has received widespread attention. However, little is known on whether people have preferences regarding tasks, and if so, whether these preferences conflict with priority instructions. Chapter 3 describes three experiments with a high-speed driving game and an auditory memory task. In Experiment 1, participants did not receive priority instructions. Participants performed different sequences of single-task and dual-task conditions. Task performance was evaluated according to participants' retrospective accounts on preferences. These preferences were reformulated as priority instructions in Experiments 2 and 3. The results showed that people differ in their preferences regarding task prioritization in an experimental setting, which can be overruled by priority instructions, but only after increased dual-task exposure. Additional measures of mental effort showed that performance tradeoffs had an impact on mental effort. The interpretation of these findings was used to explore an extension of Threaded Cognition Theory (Salvucci & Taatgen, 2008) with Hockey's (1997) Compensatory Control Model.

Chapter 4 investigates how task prioritization influences coping behavior under varying task demands. As a first step, the logics are postulated to infer the coping strategies of Hockey's (1997) Compensatory Control Model from tradeoff patterns in dual-task performance and effort, with explicit attention to task prioritization. Two new coping strategies follow from these logics, labelled as 'intense focus' and 'exclusive decrement'. The second step concerns a dual-task experiment with an auditory memory task and a driving task, based on the context of police work. Task demands were manipulated through signal-to-noise ratio and route curvature. For each of two priority instructions (driving, equal), coping strategies were inferred through pair-wise comparisons between the experimental conditions. Expected coping strategies were found in comparisons between single-task and dual-task conditions, but not in comparisons between pairs of dual-task conditions. Furthermore, none of the comparisons yielded an identical coping strategy for both instructions. Therefore, it is essential to involve task prioritization in dual-task driving studies.

Chapter 5 examines how transitions in task demand are manifested in mental workload and performance, in a dual-task setting. Hysteresis has been defined as the on-going influence of demand levels prior to a demand transition. Previous studies predominantly examined hysteretic effects in terms of performance. However, little is known about the temporal development of hysteresis in mental workload. A simulated driving task was combined with an auditory memory task. Participants were instructed to prioritize driving, or to prioritize both tasks equally. Three experimental conditions with low, high, and low task demands were constructed by manipulating the frequency of lane changing. Multiple measures of subjective mental workload were taken during experimental conditions. Contrary to our prediction, no hysteretic effects were found after the high-to-low demand transition. However, an hysteretic effect in mental workload was found within the high demand condition, which degraded toward the end of the high condition. Priority instructions were not reflected in performance. Online assessment of both performance and mental workload demonstrates the transient nature of hysteretic effects. An explanation for the observed hysteretic effect in mental workload is offered in terms of effort regulation. An informed arrival at the scene is important in safety operations, but peaks in mental workload should be avoided to prevent buildup of fatigue. Therefore, communication technologies should incorporate the historical profile of task demand.

In Chapter 6 the findings of the observations studies and experiments in the previous chapters are integrated. A shortlist is provided with key findings, and practical implications are presented for the development of information technology in police vehicles. Finally, two areas of further research are discussed. One research area concerns validation of the theoretical model developed in Chapter 3, aided by the postulated logics in Chapter 4 to infer coping strategies. The other research area concerns a new interpretation of work overload, based on workflow fragmentation (Chapter 2) and hysteresis (Chapter 5).

Samenvatting

Dit proefschrift omvat de resultaten van een reeks studies over het beluisteren van radioberichten terwijl men gelijktijdig auto rijdt. Deze taakcombinatie komt vaak voor in politiewerk, maar wordt onveilig geacht voor reguliere autobestuurders. De focus van het onderzoek ligt op taakprioritering; een proces waarin aandacht wordt toegekend aan de ene taak, maar ten koste van andere taken. In tegenstelling tot reguliere autobestuurders hebben politieagenten normaliter niet de keuze om hun voertuig te stoppen wanneer een belangrijk bericht binnenkomt. Agenten kunnen het zich echter ook niet veroorloven om ongeïnformeerd op een melding te arriveren. Met andere woorden, van reguliere autobestuurders wordt verwacht dat zij de rijtaak volledig prioriteren, terwijl van agenten verwacht wordt dat zij de rijtaak en de communicatietaak gelijkwaardig prioriteren. Deze verwachting zou een verklaring kunnen zijn voor recente berichten omtrent overbelasting bij politieagenten. Om deze reden is het doel van dit proefschrift om de mechanismen die met taakprioritering te maken hebben in een dynamisch socio-technisch systeem als de politiecontext beter te begrijpen. De opgedane inzichten vanuit de cognitieve psychologie kunnen nuttig zijn voor de ontwikkeling van informatietechnologie in toekomstige politievoertuigen.

Hoofdstuk 2 verkent hoe de noodzaak van taakprioritering afhangt van de werksituatie waarin politieagenten zich bevinden. Politiewerk kenmerkt zich door een continu schakelen tussen surveillance, noodhulp en bureauwerk. Transities tussen deze activiteiten worden doorgaans ingeleid door radioberichten, door berichten op de boordcomputer, of door persoonlijke observaties. De informatieomgeving die vanuit deze bronnen ontstaat zou een oorzaak van mentale overbelasting kunnen zijn. Hoewel de fragmentarische aard van politiewerk onderkent wordt in de vakliteratuur, zijn hierover weinig details bekend. Het doel van deze studie is daarom om vanuit de dynamiek van politiewerk een beter inzicht te verkrijgen in de cognitieve belasting van agenten. In de observatiestudie zijn tien agenten van de Nationale Politie vergezeld in hun voertuig en is 28 uur aan data verzameld. Activiteiten werden vastgelegd in vooraf opgestelde categorieën. Hierbij werd per activiteit gekeken naar hoe agenten hun informatieomgeving ervoeren. De data is vastgelegd in zogenaamde 'Transitional Journey Maps', een nieuwe methode om werk te visualiseren. In deze methode worden geobserveerde activiteiten gekoppeld aan ervaringen en contextuele informatie. De methode is gebruikt om situaties waarin zich mentale overbelasting voordeed te identificeren, alsmede om verschillen tussen solo en duo surveillance te verklaren. De inzichten die hieruit verkregen zijn zijn van belang voor het verbeteren van de informatiesystemen die agenten in hun voertuig ondersteunen.

Tevens beschrijft Hoofdstuk 2 een observatiestudie uitgevoerd in satellietcontrolekamers. Deze studie is uitgevoerd om te testen of de 'Transitional Journey Map' methode ook in andere operationele contexten toegepast kan worden. Vijf controllers in drie controlekamers zijn geobserveerd om de functionaliteit van de alarmsignalen in de controlekamers te evalueren. In totaal werden 31 alarmen geregistreerd, waarvan slechts twee alarmen aanleiding gaven tot probleemoplossend gedrag. De 'Transitional Journey Maps' zijn gebruikt om de observaties en de kwalitatieve interviewdata te bundelen. Een analyse op de uitspraken van de controllers toonde aan dat verreweg de meeste alarmen geanticipeerd waren. Hierop zijn aanbevelingen gedaan ter verbetering van het ontwerp van de alarmsignalen. Een vergelijking van beide observatiestudies toont aan dat het visualisatieproces veel overeenkomsten vertoont met het schetsproces van ontwerpers. Een belangrijke conclusie is dat het niet noodzakelijk is om vooraf categorieën te definiëren, mits men voldoende data vergaart om deze op verschillende abstractieniveaus te kunnen beschrijven.

Er is weinig bekend over of men voorkeuren heeft in taakprioritering, noch over de mogelijke invloed van deze voorkeuren op het volgen van instructies met betrekking tot taakprioritering. Hiertoe zijn in Hoofdstuk 3 een drietal experimenten uitgevoerd waarin een rij-spel is gecombineerd met een auditieve geheugentaak. In Experiment 1 kregen participanten geen prioriteringsinstructie. De participanten voerden in verschillende volgordes enkele of gecombineerde taken uit. Na afloop werd hen gevraagd aan welke taak het meeste aandacht was gegeven. Deze aandachtsverdeling is geïnterpreteerd als preferentie en verschillende preferenties bleken als zodanig terug te vinden te zijn in de taakprestaties. In Experimenten 2 en 3 zijn de formuleringen van de preferenties gebruikt om prioriteringsinstructies op te stellen. De resultaten tonen aan dat mensen verschillende preferenties hebben, welke overheerst kunnen worden door instructies, maar alleen na voldoende ervaring met het gelijktijdig uitvoeren van beide taken. Verder is een wisselwerking geconstateerd tussen taakprestatie en mentale belasting. De interpretatie van bovenstaande bevindingen heeft aanleiding gegeven tot het verkennen van een integratie van twee bestaande theorieën, te weten 'Threaded Cognition Theory' (Salvucci & Taatgen, 2008) en het 'Compensatory Control Model' (Hockey, 1997).

In Hoofdstuk 4 wordt onderzocht hoe taakprioritering beïnvloedt welke strategieën met hanteert bij het omgaan met taakbelasting. Hockey's (1997) 'Compensatory Control Model' voorspelt een verzameling van strategieën. Als eerste stap is op basis van logica een aantal regels opgesteld van waaruit men deze strategieën kan herleiden. De ingrediënten voor deze regels zijn taakprestatie, mentale belasting en taakprioritering. Op basis van dezelfde logica blijkt dat Hockey’s verzameling aangevuld zou kunnen worden met een tweetal nieuwe strategieën. Als tweede stap is een experiment uitgevoerd met een rij-spel en een auditieve geheugentaak. Taakbelasting is gemanipuleerd door enerzijds de signaal-ruis verhouding te variëren en anderzijds de scherpte van de bochten in de te rijden routes. De te verwachten strategieën zijn gevonden in overgangen van condities met een enkele taak naar condities met beide taken gecombineerd. Echter, geen van de te verwachten strategieën is gevonden in overgangen tussen condities met beide taken gecombineerd. Verder bleken de prioriteringsinstructies in geen enkele vergelijking te leiden tot eenzelfde strategie. Deze laatste bevinding toont aan dat het essentieel is om taakprioritering te betrekken in onderzoek naar simultane taakprestatie.

Het doel van de studie in Hoofdstuk 5 was om te verkennen hoe een plotselinge overgang in taakbelasting in de daarop volgende tijd een effect heeft op mentale belasting. Dit effect wordt ook wel aangeduid als ‘hysterese’. Wederom is een auditieve geheugentaak gecombineerd met een rijtaak, al vond deze ditmaal in een rijsimulator plaats. Participanten moesten ofwel de rijtaak prioriteren, ofwel beide taken gelijk prioriteren. De belasting van de rijtaak is gemanipuleerd door de afstand tussen te ontwijken obstakels te variëren. Hierdoor ontstonden drie experimentele condities, met lage, hoge en wederom lage taakbelasting. Gedurende de ritten werd meerdere keren gevraagd hoe zwaar men de taak ervoer. In tegenstelling tot wat verwacht werd, is geen hysterese geconstateerd bij de overgang van hoge naar lage taakbelasting. Wel is er hysterese gevonden in mentale belasting bij de eerste overgang, van laag naar hoog. Het hysteretische effect verminderde naarmate het einde van de conditie met hoge taakbelasting werd bereikt. Hieruit blijkt de noodzaak om herhaaldelijke metingen uit te voeren binnen de experimentele condities. De geconstateerde hysterese kon worden verklaard vanuit bestaande theorieën over het reguleren van mentale belasting.

Hoofdstuk 6 integreert de bevindingen uit de observatiestudies en experimenten van de eerdere hoofdstukken. De hoofdbevindingen zijn samengevat, alsmede praktische aanbevelingen voor de ontwikkeling van informatietechnologie in toekomstige politievoertuigen. Tenslotte zijn twee nieuwe onderzoeksgebieden gepresenteerd. Het eerste onderzoeksgebied betreft het valideren van het theoretische model uit Hoofdstuk 3. Deze validatie kan ondersteund worden met de logica uit Hoofdstuk 4 aangaande het herleiden van strategieën. Het tweede onderzoeksgebied betreft een nieuwe interpretatie van het fenomeen overbelasting, gebaseerd op werkfragmentatie (Hoofdstuk 2) en hysterese (Hoofdstuk 5).

Acknowledgments

At the end of this thesis, I would like to thank the people who have helped me along this journey, starting with my supervisors.

Huib, you continue to amaze me with connecting the dots of larger pictures that others fail to see. Your vision and guidance throughout my project has been of immeasurable value. Moreover, your critical reflections have strengthened my skills in logical reasoning. I will miss our sparring sessions, both in Delft and in Oegstgeest. René, thank you for giving me the space to develop myself as an independent researcher in the best possible way. I have worked for and with you for ten years. It is impossible to summarize all memories in this short space, but listening to The Beatles while driving to Drachten was always a pleasure, and our trip to Utah is definitely among my most cherished moments. If only teaching could always be like this!

A large part of this project was carried out at the request of the Dutch National Police. Therefore, I am greatly indebted to Mr. Tanis for giving me the opportunity to perform a research project within the police context. Many thanks also to Sasha, for making the project concrete and for our many conversations on work and drums while driving between Delft and Driebergen.

Thank you people of ID-Studiolab for offering a great place to blossom as a researcher, and for sharing the good, the bad, and the inevitable experiences of doing a PhD project. You have endured almost a decade of annoying product sounds and many requests to play driving games, yet still awarded me with an incredibly warm goodbye party. I will return on some unexpected day around twelve o'clock to announce lunchtime. A special thanks to Rob, after so many talks about playing bass guitar the lab was never quite the same after you left. And Elif, where would I be without your shared enthusiasm for (product) sound design? Your lectures made me a better teacher. Through ESA you made it possible for me to do some more research on sound after all. One day we will bundle our forces again...

Thank you Bryan and Karen for your hospitality, and for the opportunity to teach product sound design to your students. You have inspired me with the value of teaching small groups, and what it means to truly know your students. Our stay in Utah is among my personal highlights of the past five years.

Thank you professor Hancock, I never could have imagined what a strike on an Italian airport could result in. I have learned so much from running the simulator experiment, and it is with great pleasure that I think back of the early morning walks and talks at the UCF campus. Thanks Ben, you are one of the few PhD candidates with whom I could share my entire thought process. We should still plan that meeting, for I am sure we will come up with a paper. Also thank you Emily and Jessica, for without your tremendous efforts our research would not have been possible.

Thank you colleagues at SWOV, and especially Nicole, Michiel, and Rob, for supporting me in finishing my PhD. The last six months would have been even more demanding without your understanding.

Jos, Betty, and André, I have been away from home more than I should have been. Your everlasting encouragement for my PhD project, as well as everything else you have done for us (and that is a lot!) means the world to

me. Thanks Gyán, for redirecting my attention to the important things in life: good food, likewise drinks, photography, recording, tinkering, and jamming. Spock, Isha, Djenghis, James, and Murphy, you are still not allowed to sleep on our bed, but I will no longer lock myself up in the writing room. And finally to my Judith. To start a PhD project is to believe in tomorrow. You have always believed in me, and tomorrow has finally become reality. From this day on, Fridays will forever be ours!

About the author

Reinier Johan Jansen (born March 9th, 1983 in Leiden) received his secondary education (VWO) at the Rijnlands Lyceum in Oegstgeest from 1994 to 2000. After a year of studying Electrical Engineering at Delft University of Technology, he enrolled in the bacholor program of Industrial Design Engineering at the same university. Specializing in sound design and auditory perception, he received his Masters' title (Design for Interaction) cum laude in 2009 on a product sound sketching tool. He worked as designer on the embodiment of a laparoscopic training environment, before applying for a grant from the Dutch National Police to start his PhD research in 2012. As part of his research, Reinier conducted a simulator experiment at the University of Central Florida, USA, under supervision of Professor Peter Hancock. Apart from his research activities, Reinier has been involved in teaching and coordinating courses on product sound design, cognitive ergonomics, and research methodology. This resulted in giving an invited workshop on product sound design at Brigham Young University, USA. In May 2016 Reinier started working as researcher at SWOV Institute for Road Safety Research, where he is involved in the analysis of naturalistic driving data. When he has free time from his research activities, he likes to play drums, bass guitar, and take photos. Reinier is happily married to Judith and they dream of many more camper van trips to Norway.

Publications

Journals

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2016). Task Prioritization in Dual-Tasking: Instructions versus Preferences. *PLoS ONE 11*(7): e0158511.

Jansen, R.J., Sawyer, B.D., van Egmond, R., de Ridder, H. & Hancock, P.A. (2016). Hysteresis in Mental Workload and Task Performance: The Influence of Demand Transitions and Task Prioritization. *Human Factors 58*(8): 1143-1157.

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2015). No Alarms and no Surprises: How Qualitative Data Informs Bayesian Inference of Anticipated Alarm Sounds. *Procedia Manufacturing 3*, pp. 1750-1757.

Jansen, R.J., Özcan, E. & Van Egmond, R. (2011). PSST! Product Sound Sketching Tool. *Journal of the Audio Engineering Society 59*(6), pp. 396-403, June 2011.

Book chapters

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2016). Transitional Journey Maps: Reflections on creating workflow visualizations. In: Arezes, P.M. & Rodrigues de Carvalho, P.V. (Eds.) *Ergonomics and Human Factors in Safety Management*. CRC Press: 271-304.

Langeveld, L., Van Egmond, R., **Jansen, R.J.** & Özcan, E. (2013). Product Sound Design: Intentional and Consequential Sounds. In: Coelho, D. (Ed.) *Advances in Industrial Design Engineering*, pp. 47-73, Intech.

Conferences

Sousa, B., Donati, A., Özcan, E., Van Egmond, R., Edworthy, J., **Jansen, R.J.**, Kristensen, M., Peldszus, R. & Voumard, Y. (2016). Designing and deploying meaningful audio alarms for control systems. *14th Intl. Conf. on Space Operations*, 16-20 May 2016, Daejeon, Republic of Korea.

Heydra, C.G., **Jansen, R.J.** & Van Egmond, R. (2014). Auditory signal design for automatic number plate recognition systems. *Chi Sparks 2014*, pp. 19-23. 3 April 2014, The Hague, The Netherlands.

Jansen, R.J., Van Egmond, R., De Ridder, H. & Silvester, S. (2014). Transitional Journey Maps: Capturing the dynamics of operational policing. *Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference*, 16-18 October 2013, Turin, Italy.

Langeveld, L., **Jansen, R.J.** & Van Egmond, R. (2012). Product Sound Design in Education. *Engineering & Product Design Education*, 6-7 September 2012, Antwerp, Belgium.

Jansen, R.J., Van Egmond, R. & De Ridder, H. (2012). Sound meets Image: Freedom of Expression in Texture Description. *Human Vision and Electronic Imaging XVII*, 23-26 January 2012, Burlingame, California.

Jansen, R.J., Özcan, E. & Van Egmond, R. (2010). PSST! Product Sound Sketching Tool. *AES 38th International Conference on Sound Quality Evaluation*, 13-15 June 2010, Piteå, Sweden.

Technical reports

Edworthy, J., Özcan, E., Van Egmond, R. & **Jansen, R.J.** (2014). *Alarm Philosophy for ESOC Mission Control Rooms in Darmstadt*. University of Plymouth, Cognition Institute.

Jansen, R.J., Van Egmond, R., Özcan, E. & Edworthy, J. (2014). *Observational Study of ESOC Control Rooms in Darmstadt*. Delft University of Technology.

Jansen, R.J., Silvester, S., Van Egmond, R. & De Ridder, H. (2014). *Eindrapport Mobiel Werken*. Delft University of Technology.