Developing a bicyclist hazard perception test:
Explorative research comparing adult and adolescent cyclists on a visual scanning and a key press measure

Name: Marnix Alexander Knoll (s1132997)
Supervisor: Drs. W.L.G. Verschuur
Second reader: Dr. J. Groeneweg

Cognitive Psychology
Thesis Master of Science Applied Cognitive Psychology
Universiteit Leiden, 2016
Abstract

Recent years show an increase in the amount of accidents resulting in serious injuries involving bicyclists. Adolescents are overrepresented in these figures. In order to gain more understanding into adolescent cyclist behavior, a hazard perception test is conducted. The independent within-group variable is distraction and the independent between group variable is experience. The dependent variables were a key press measure (having a button pressed whilst the latent hazard is present) and an eye tracking measure (looking at the latent hazard for 200ms or longer). Using real traffic videos from the viewpoint of a cyclists, a computer test consisting of 22 items was developed for this experiment. The experiment shows that adults performed significantly better than adolescents on the key press score, but no significant difference is found for eye tracking. In conclusion, adults outperform adolescents into indicating there is a danger, but not in spotting dangers. Future research should prefer the key press measure and focus on improving the psychometric qualities of bicyclist hazard perception tests.

Keywords: adolescent cyclists, hazard perception, eye tracking, test development, key press measure
Table of contents

1. Introduction
   1.1 Relevance of this research ............................................. 5
   1.2 Aim of this research ..................................................... 7
   1.3 Hazard perception ....................................................... 8
   1.4 Theoretical frameworks of HP: situation awareness .............. 10
   1.5 Theoretical frameworks of HP: information processing strategies ...... 11
   1.6 Theoretical frameworks of HP: workload .......................... 13
   1.7 Theoretical frameworks of HP: bicyclist hazard perception ........... 16
   1.8 Purpose statement ...................................................... 16
   1.9 Research questions .................................................... 16

2. Method
   2.1 Ethics ........................................................................... 19
   2.2 Participants ................................................................. 19
   2.3 Design .......................................................................... 19
   2.4 Motivation for dependent variables ..................................... 20
   2.5 Stimulus material .......................................................... 20
   2.6 Apparatus ..................................................................... 21
   2.7 Secondary task .............................................................. 22
   2.8 Procedure ...................................................................... 23
   2.9 Analysis ........................................................................ 25

3. Results
   3.1 Results of the experiment ................................................ 27
   3.2 Additional results ........................................................... 29
   3.3 Results surrounding test development: Key press................... 31
3.4 Results surrounding test development: Fixations..........................32

4. Discussion

4.1 The experiment................................................................. 35
4.2 Test development............................................................ 37
4.3 Conclusion........................................................................ 39

5. Limitations and recommendations

5.1 Limitations....................................................................... 41
5.2 Recommendations............................................................ 41

6. References............................................................................. 43

Appendix A: Descriptions of the stimulus material............................. 46
Introduction

The Netherlands has seen a reduction in traffic accidents and deaths among all traffic participants between 1970 and 2006, due to improvements based on vehicle design, knowledge of infrastructure and insight into road user behavior (Stichting Wetenschappelijk Onderzoek Verkeersveiligheid [SWOV], 2007). In contrast to this trend, the number of road users who get severely wounded in traffic has increased in the years after 2006. Of those who get severely wounded in traffic, half are cyclists. Their risk of accidents resulting in serious injuries per distance traveled had increased between 2006 and 2009 (SWOV, 2013a). Of those who died in traffic in 2015, over one third were cyclists (SWOV, 2016). The number of cyclist deaths has been stable over the past years, with 184 deaths in 2013, 185 in 2014 and 185 in 2015 respectively (Centraal Bureau voor de Statistiek [CBS], 2016b). There is a group overrepresented in these traffic accidents statistics, adolescents (Twisk, Bos, Shope & Kok, 2013).

Among adolescents aged 10-17, traffic accidents are one of the main causes of death (Twisk, Bos, Shope & Kok, 2013). People aged 12 to 24 form just 16% of the population in the Netherlands in 2015 (CBS, 2016a), yet they are involved in around 33% of bicycle related accidents (Twisk, Bos, Shope & Kok, 2013).

The adolescent bicyclist group has some unique characteristics that are brought forward as potential causes for accidents. These can roughly be summed up on three categories. Firstly inexperience, they have less years of cycling experience than adults. Secondly age factors associated with their developmental state, such as novelty seeking and risk taking (Steinberg, 2004). And thirdly, making the most kilometers on bike compared to other age groups. They are known for using their bike as their primary mode of transport. That the bike is a preferred method of transport can be explained by the Netherlands being a late licensing country (age 18 for car and motorcycle). This paper will focus on comparing the low experience and high risk group of adolescents to the higher experience and lower risk group of adults. But before we got to the experiment, let’s look at these three factors that have been put forward as explanations for the increase in risk more closely.

Firstly, the factor inexperience is argued to be the cause of increased risk for traffic accidents. Experience can be defined as years of cycling experience. Adolescents start cycling to their secondary school around age 12, generally exposing them to larger travel distances
than during elementary school and exposing them to new traffic situations. They can be considered to be low in cycling experience, considering elementary school students (4-11) generally travel short distances to their school, are often guided by their parents and they are more risk averse in traffic (Oron-Gilad, Meir, Tapiro & Borowsky, 2011). Early cycling experience has been shown to be a protective factor (Twisk, Bos, Shope & Kok, 2013). Consequently, those students without cycling experience during elementary school and thus without this protective factor run a higher risk for accidents. In line with the hypothesis that less experience leads to higher risk, Chinn, Alliot, Sentinella and Williams (2004) report the number of serious injuries and deaths in traffic peaks in the first few years of secondary school.

Secondly it is shown that adolescents are more risk taking than adults, even on the road (Arnett, 1994). Youngsters (12-24) have significantly more accidents with racing/stuntin as a cause compared to adults. This could still be considered a small factor, considering racing/stunting makes up only 6% of accidents of youngsters compared to 0% for adults (24 years old or over). The cause of the risk-taking behavior is unclear, but multiple lines of reasoning have been put forward to explain it. It could be that adults and adolescent perceive the same risks, yet adolescents are more risk accepting. Another line of reasoning is that adolescents do not perceive the same risks, and thus show more sensation seeking behavior due to poor perception and risk judgements. This experiment will look at eye scanning patterns precisely to get a clearer image which line of argument holds more ground when compared to evidence. What is already known is that sensation seeking peeks around age 18,5 for boys and around 16 for girls (Romer & Hennessy, 2007), yet statistics suggest the accidents peak during the first years of secondary school (age 12-14). This result points toward the conclusion that risk taking likely isn’t the explanation for increased risk of bicyclist accidents during the first years of secondary school. Yet, it doesn’t point to any conclusion on differences between adults and adolescents in perception.

Thirdly the risk per distance traveled is a good indicator of how groups compare in traffic safety. SWOV (2013a) reads that teenagers show no clear increase of risk per distance traveled on bike compared to adults. The age group of 16-17 shows no increase in risk of severe injuries compared to 12-15 year olds and 18-54 year olds. This data points to the higher number of accidents being caused by the higher distance traveled by the adolescents and not by their increased risk. Consequently, it points to the possibility that no significant
differences in behavior can be expected between adolescents and adult bicyclists, since no significant differences in risk are reported.

**Aim of this research**

The purpose statement and research questions will follow after the theoretical framework. This section serves as a broad oversight of the research project. This research project has three main aims.

The first aim is to analyze the factor experience by comparing adolescents (low experience, high accident risk group) and adults (experience group, lower accident risk) on the measures of hazard perception. Hazard perception is defined as the ability to detect developing, potentially hazardous, traffic situations. Hazard perception will be measured in the form of gazing locations (does the participant see the developing danger) and indications of a heightened feeling of developing danger (does the participant feel that the developing situation requires extra attention). The first aim is to detect differences in hazard perception due to experience.

The second aim is to analyze if the theories used to describe hazard perception among car drivers, can be used to explain and predict the results of hazard perception among cyclists.

The third aim is to set a first step in the development of a cyclist hazard perception test that meets both the requirements for validity and reliability. Due to its premier nature, a closer analysis of what makes stimuli good discriminators and what makes the test valid is required.

The broad questions that will be explored to answer these three aims are:

- Do differences exist between high accident risk cyclists and low accident risk cyclists on the cognitive ability of hazard perception?
- Does the current theory explain the relationship between hazard perception and experience or accident risk among bicyclists?
- What is the influence of the stimulus type on their strength in distinguishing groups? What can be done to make future cyclist hazard perception tests better equipped to distinguish between low and high skilled cyclists on the ability of hazard perception?
Hazard perception

Hazard perception [HP] is a cognitive ability (or skill) which is related to traffic safety and in which age and experience have been shown to be significant (predictive) factors. Hazard perception has been defined as ‘the ability to predict dangerous situations on the road’ (Wetton, Hill & Horswil, 2013). Vlakveld (2014) offers a more extensive definition: ‘the ability to detect and recognize possible hazard and to predict how these possible hazards can develop into situation in which a crash would be very likely. Vlakveld (2014) his definition links hazard perception to situational awareness and will be the accepted definition in this paper.

Hazard anticipation is a more encompassing term. Hazard anticipation is a subject closely linked to hazard perception. It represents both the hazard perception and the consequent action taken. Since this research project will also focus on the actions or responses of participants to latent hazard, hazard anticipation will be used when describing the action part in addition to the perception part.

Hazard perception has been shown to be a measurable skill that can be used to predict accident risk among car drivers (Horswill, Hill & Wetton, 2015). People skilled at hazard perception are better at detecting cues that predict hazardous situations and knowing (and sensing) these cues are predictors for hazardous situations compared to others with a poorer hazard perception skill. For hazard anticipation, skilled people are also better at choosing the (correct) actions that will put them at a significantly lower risk for car accidents, compared to others with a poorer hazard anticipation skill.

A hazard (a risk or a danger) in this thesis refers to traffic situations involving other (potentially hidden) road users which may interfere with the road users goal of arriving at the destination safely. It’s important to differentiate between acute and latent hazards. An acute hazard is one to which you can only respond by reflex. In contrast, latent hazards are hazards that can potentially manifest themselves. Unlike reflexes for the acute hazards, latent hazards require a more predictive or anticipatory behavior in order to deal with them.

Within the framework of hazard perception, the hazard is always latent. This is important considering responses to acute threats do not give insight into the road user’s ability to predict how traffic situations will develop, but rather represents reaction speed. The goal is to measure the predictive skill, hazard anticipation, the ability to correctly judge how a situation could (or will) play out. This is most comparable to real life traffic situations in
which you can never perfectly predict how other road users will behave, since you only control your own behavior, and in which you have to make the best judgement with the information you have. This predictive ability lies at the core of hazard perception and hazard anticipation.

Latent hazards can be either overt or covert (Vlakveld, 2014). Overt refers to other road users that are visible. An overt latent hazard is thus a visible other road user that might cause you to brake or change in direction in order to safely arrive at your destination. Examples are pedestrians, bicyclists or cars that could possibly end up on a collision course with your vehicle or person. Covert on the other hand refers to possibly hidden road users. These are potential road users that might be hidden from view by a physical blockade of your visual field. Examples of these blockades are walls, trees, fences, other road users, parked cars or even sporadic objects like trashcans or bus stops. Behind these objects there can be another road user which you can’t perceive, but which may be there and consequently may end up on a potential collision course with you. A covert latent hazard is thus a potential road user hidden from view by a vision blockading object and not yet overt, that might possibly end up on a collision course with your vehicle or person. Examples of covert hazards are alleyways where road users may suddenly come out, parked cars hiding children that might run on the street or trucks taking a turn hiding another car passing behind them.

The factor experience positively impacts hazard perception. Research has shown that novice drivers perform worse on hazard perception than experienced drivers, who in turn perform worse than police officers (Crundall, Chapman, Phelps & Underwood, 2003; Horswill, Taylor, Newnam, Wetton & Hill, 2013). Since hazard perception has been shown to be a measurable skill that can used to predict accident risk among car drivers, the assumption can be made that experience leads to a better hazard perception skill which leads to a lower accident risk (Horswill, Hill & Wetton, 2015).

Herslund and Jørgensen (2003) describe the phenomenon ‘looked-but-failed-to-see’ in their paper. This phenomenon occurs when a road user reports not having seen another road user until immediately before a collision, even though this other road user was clearly visible. Within hazard perception, it’s thus important to check not only if people looked at the other road user, but also whether they perceived this road user.

Taking into account hazard perception as a relevant factor when thinking about accident risk and road safety has shown to be beneficial in improving safe behaviors of car
users. Consequently, hazard perception could also turn out to be beneficial for the bicyclist population. In contrast to the wealth of research on this factor in cars operators, it has not been widely researched among bicyclists. To correctly understand how to replicate this research within another vehicle population, it is imperative to have insight into the theoretical frameworks underlying hazard perception in car drivers.

Different theories have been used to analyze the question of why people perform at hazard perception tests the way they do or even to analyze what hazard perception is. In line with the tradition, this paper will analyze multiple theories and their implications to form an informed starting point from which to research hazard perception and hazard anticipation among bicyclists. The theories that will be looked at more closely are situation awareness (Endsley, 1995), information processing strategies (Wickens et al., 2001) and workload theory (DiDomenico & Nussbaum, 2008).

**Theoretical frameworks of HP: situation awareness**

Situational awareness is a theoretical framework used to describe how people form and update their mental model of the current environment as it relates to their current goals. This is a broader theoretical framework than hazard awareness and can be used as a basis for analyzing perception in a dynamic setting such as traffic.

Jeannot, Kelly and Thompson (2003) describe situation awareness as a four-element linear process that forms a constant loop which makes up our situational awareness. This linear process starts with (1) extracting information from the environment, (2) integrating the information with relevant internal knowledge to create a mental picture of the current situation, (3) using this picture to direct further perceptual exploration in a continual perceptual cycle and (4) anticipating future events.

Another popular comparable model is the three-level model of situational awareness by Endsley (1995). Here the linear process starts with (1) perception of elements or cues in the environment, (2) comprehension of their qualities and relevance to current goals and (3) projection of their locations in space over a suitable timeframe for the task at hand (Bolstad, Cuevas, Wang-Costello & Endsley, 2010; Crundall, 2016).

Hazard perception has been studied from a situational awareness point of view, primarily with research using the model by Endsley (1995). With hazard perception, the linear process also starts with (1) perception of elements or cues in the environment, (2) comprehension of their qualities and relevance to current goals and end with (3) projection of
their locations in space over a suitable timeframe for the task at hand. For hazard perception respectively these steps are (1) perceiving the cue or potential hazard, (2) understanding (and feeling) the risk associated with the potential hazard and (3) predicting the upcoming future event (Crundall, 2016).

In contrast to the standard of using the model by Endsley (1995), one could argue the model by Jeannot, Kelly and Thompson (2003) should be preferred when thinking about hazard perception. The key difference is that the first model by Jeannot, Kelly and Thompson (2003) puts more emphasis on using earlier cues as a motivator for perceptual exploration later on, placing the mental schema that one created of a situation in a ‘bigger picture’ that moves through time and influences the next perceptual exploration. This is in line with the real life situation whilst bicycling. Contrasting to this, the second model represents more a picture than a movie, looking more closely at what a stimulus means to an individual rather than how it influences what stimulus you will focus on next. Precisely for the reason this represents reality more closely, the model by Jeannot, Kelly and Thompson (2003) should be preferred.

**Theoretical frameworks of HP: information processing strategies**

Information processing strategies have also focused on the function of stimuli when forming a perception, but in contrast to the situational awareness frameworks they focus on explaining visual search and attention allocation (step 1 of situation awareness). The SEEV model by Wickens et al. (2001) explains visual search and attention allocation in dynamic environments using a formula. The probability that a given area will be attended (P(A)) is determined by the salience (S) of an event minus effort (E) plus expectancy (E) times value (V).

\[
\text{Probability ( Attended)} = \text{Salience} - \text{Effort} + (\text{Expectancy})(\text{Value})
\]

Salience is determined by the physical properties of an event. Effort is determined by the physical distance between a previously fixated event and the current event or effort is determined by the demands of concurrent tasks. Expectancy is determined by what is to be expected in this situation based on information or knowledge one has. Value is either determined by the value someone ascribes to getting the event or determined by the cost of
Werneke and Vollrath (2012) perceived this formula to fit in with two information processing strategies: top-down and bottom-up. Their model also flows from the assumption that visual attention is guided by four factors: Salience, Effort, Expectancy and Value (SEEV). Salience and effort are grouped as passive, bottom-up, and group expectancy and value are grouped as top down. This results in the model seen in figure 1.

![Figure 1. The SEEV model explaining visual search and attention allocation. (adapted from Werneke & Vollrath, 2012).](image)

Bottom-up and top-down processing strategies are related to situational awareness in that they also explain how perceptions are formed. After all, an individual will only visually perceive that which he grants visual attention. From a cognitive psychology perspective, bottom up refers to an information processing strategy in which sensory input forms the basis when forming a perception. Fast reactions or quick visual identification are trademark bottom-up processes, driven by sensory information. In contrast the top down information processing strategy reads that perception is an active and constructive process. Directed attention is considered top-down, because the direction of the attention is formed based on goal direction and the individual rather than the sensory input. The second, third and fourth step in in Jeannot, Kelly and Thompson (2003) their model can be seen as top-down processes. The situational awareness is formed by both bottom-up and top-down processes.
For both processing strategies examples can be found within the hazard perception literature. An example of bottom-up processing is that ‘highly visible road users are less likely to be involved in crashes, suggesting that saliency is important in real-world tasks’ and ‘high saliency can attract early fixation’ (Underwood, Humphrey & van Loon, 2011). Top-down processing shows up in hazard perception when eye moments show that attention is directed towards locations that don’t have a salient stimulus, but might produce a hazard, thus being driven by value and expectancy forming a top-down visual strategy. Directed attention of experienced drivers often focuses more on covert latent hazards and environmental prediction compared to novice drivers (Borowsky, Shinar & Oron-Gilad, 2010). This difference suggests that these groups use different top-down tactics in their visual search, with experienced drivers expecting traffic from covert locations, giving more value to covert locations or any mix of the two (Crundall, Chapman, Trawley, Collins, van Loon, Andrews & Underwood, 2012; Crundall, 2016). Information processing strategies could be a promising starting point from which to analyze hazard perception.

For this paper, information processing strategies aid in explaining the link between visual searching patterns and hazard perception. Two main points can be drawn from the theory. Firstly, based on the SEEV model (Werneke & Vollrath, 2012), a link can be made between fixations and visual attention allocation. When participants fixate on an area than it is probable it has been attended or given attention. Secondly the importance of top-down processes is stressed. When fixations are not at the most salient area, it likely represents a combination of expectancy and value based on a top-down model of the participant. Within hazard perception, experience could lead to a more refined top-down model that draws attention to less salient, but nonetheless more important for traffic safety, areas of the visual field. Thus, it is reasonable to assume that fixations on covert hazards could point towards hazard perception. For overt hazards, it is harder to distinguish between saliency on one side and top-down guided searching on the other.

**Theoretical frameworks of HP: workload**

In this research project, the influence of workload has been counter balanced among conditions and will thus only be described to give a full picture of the theory surrounding situational awareness and HP. The master thesis of the colleague I conducted this research project with (de Geus, 2016) focuses more on the influence of workload and secondary tasks. For more information on this subject I suggest the reader to read her master thesis. In short, de
Geus (2016) shows that there is no significant difference between the influence of a secondary task on adults and on adolescents.

Gugerty (2011) places situational awareness in the context of cognitive and perceptual resources. He describes three levels of cognitive processing. Firstly, there are automatic, pre-attentive processes. These occur unconsciously and place almost no demand on cognitive resources. Secondly there are recognition-primed decision processes. Here a stimulus reaches consciousness for a brief period of time and places few demand on cognitive resources. Thirdly, there are the conscious, controlled processes that place heavy demands on cognitive resources. This links in closely to the theory surrounding workload, which analysis the role cognitive resources play in task performance.

Based on the theory surrounding workload, the assumption is that the level of performance suffers as a task demands more cognitive and perceptual resources. DiDomenico and Nussbaum (2008) refer to workload as the cost incurred by the individual to achieve a particular level of performance on a certain task. Additionally, they described workload as “determined by the interaction of the task demands, the circumstances under which it is performed, and the skills, behaviors, and perceptions of the individual.” Workload is the result of multiple factors interacting. The task demands may be physical and mental in nature and occur at the same time. These demands do not determine the workload in themselves. DiDomenico and Nussbaum (2008) explain “the impact of these demands is, in turn, dependent on the abilities of the individual performing the task.” The costs incurred by the individual are thus dependent not only on the task, but also on the skill level of the individual. A skilled individual would require less cognitive resources to reach a particular level of performance than an unskilled counterpart.

The simplest way to prove the influence of workload on performance is to add a secondary task that places demands on the resources required for the first task. As a consequence of the demands of the second task, there are less resources available for the first task. This will lead to a negative effect on performance on the first task. If tasks are automatic the first task will be influenced little to none by the additional task, especially when comparing these tasks to conscious controlled tasks. Figure 2 shows the expected negative effect of an additional task on the performance of the original task of driving.
Drivers often have to perform complex visual/cognitive tasks to safely navigate the roads. The individual has to incur certain costs or efforts to achieve a particular level of performance on the task of driving. This workload is dependent on the task demands, the circumstances under which it is performed and the skills, behaviors and perceptions of the individual. The task demands during driving are both physical and mental in nature. The individual has to operate the vehicle and be aware of his surroundings to safely navigate traffic. Training would make the driver more skilled and thus lower the demand the task places on their cognitive resources or improving level of performance.

Hazard perception is one of the sub-tasks of driving that demands cognitive resources and a task that never becomes fully automatic. This can be concluded from multiple papers in which hazard perception is tested, which do not require participants to do any other tasks (i.e. operate the vehicle), yet they still find possible effects of workload and secondary tasks on participants their hazard perception skill. For example, McKenna and Farrand (1999) show that hazard perception is negatively influenced by a secondary task in both experienced and novice drivers. Multiple papers support the assumption that hazard perception and general visual search suffer under added task load. Mackenzie and Harris (2015) show that drivers are slower at detecting hazards whilst driving compared to a more passive video-based hazard perception task. Solman, Cheyne and Smilek (2011) show that memory load effects visual search processes in general. Matsukura, Brockmole, Boot and Henderson (2011) show that a
dual task reduces the likelihood and speed of oculomotor capture. Harbluk, Noy and Eizenman (2002) found that increased cognitive load led to more staring behavior. The cognitive load led participants to make fewer saccades, spend more time looking centrally and spend less time looking into the periphery. In addition, they found that the increase in cognitive load led to a decrease in driving performance in the form of more incidents of hard breaking.

**Theoretical frameworks of HP: bicyclist hazard perception**

Considering this research project focuses on the differences between a high cycling experience group with lower accident risk and low experience group with a higher accident risk, with a counterbalanced secondary task, information processing strategies put forward the most useful theoretical influence on this project. Using this framework, the expectation is that low experience cyclists are significantly worse at hazard perception compared to more experienced cyclists. Information processing strategies describe that the basis of the difference between these groups lies within top-down visual attention location. In line with this expectation, this research project uses multiple covert hazards that distinguish top-down visual search from stimulus driven visual search. From this theoretical basis, the purpose statement and research questions follow.

**Purpose statement**

The intent of this study is to discover the difference of hazard perception, a higher cognitive ability, between low accident risk cyclists, adults, and high accident risk cyclists, adolescents. Hazard perception is defined as the ability to detect developing, potentially hazardous, traffic situations. In multiple closed off research rooms, a computer test will be conducted to determine hazard perception in the form of gazing locations (does the participant see the developing danger) and indications of a heightened feeling of developing danger (does the participant feel that the developing situation requires extra attention). The purpose is to see if the grounded theory from hazard perception among car driver, holds among the cyclist populations. In addition, this test developed for this study will be analyzed more closely in order to discover potential learning points of how to develop future hazard perception tests for cyclist populations.

**Research questions**

Due to its proven effectiveness among drivers and strong basis in literature, hazard perception potentially has a lot of insights to offer into cyclist behavior and its relation to accident risk. Therefore, this study will explore the possibility to distinguish high accident
risk cyclists from low accident risk cyclists by the measurement of two forms of hazard perception, indicating hazard and perceiving hazards respectively. In addition, this study aims to question what items ought to be used when testing cyclist hazard perception. The research questions are the following:

1. Do differences exist between high accident risk cyclists and low accident risk cyclists on the higher cognitive ability of hazard perception? What are the effects on experience on hazard perception? If differences do occur, do they show the expected direction?
   a. What are the differences between the scanning or ‘seeing dangers’ scores of the high accident risk and low accident risk group?
   b. What are the differences between the feeling danger scores of the high accident risk and low accident risk group?

2. Does the information procession theory explain the relationship between hazard perception scores and accident risk among bicyclists, controlling for the effects of distraction?
   a. Do the scores on hazard perception show the same differences found in previous research among other vehicle types?

3. Which effect does distraction have on hazard perception? Is the effect the same for different risk groups?

4. Does the test manage to distinguish the different experience groups better than chance? Which (kind of) stimuli show to be the most promising in distinguishing low and high risk cyclists? Is the test reliable in the form of internal consistency?
   a. What is the discriminant validity of the test and is it better than random chance?
   b. What is the value of different stimulus types in distinguishing groups on hazard perception for low accident and high accident risk cyclists?

These questions should help to assess what role hazard perception plays in explaining cyclist behavior and accident risk. If, for example, it shows to be an effective method in distinguishing low risk and high risk cyclists, hazard perception ought to deserve more attention in plans to reduce accident risk. In contrast, if hazard perception shows to be of little to no significant meaning, other roads in inquiry ought to be explored.
Methods

Ethics

In order to ensure the ethics of this research, an independent committee was asked to review the test protocol. The test protocol of this research project was approved by the ethical committee of SWOV. Additionally, both the parents of the adolescent and the adolescent signed the informed consent to participate in the project.

Participants

The test was completed by (low experience) adolescents and (experienced) adult bicyclists who were randomly assigned to their respective condition. The requirements for the lower experience group were to be in the first year of secondary school and use their bicycle as a vehicle or transport to school. The requirement for experienced cyclists was to bike regularly (as specified by two or more days a week with a minimal distance of two kilometers a day) and not be 18 or under.

The adolescents were recruited in cooperation with a secondary school in the Netherlands. Teachers at this school handed out a request for pupils in the first year of their secondary education to participate in our experiment. The adult participants were recruited from multiple sources. First from companies in the vicinity of SWOV research institute, one of the primary test locations. Additionally, one adult participant was recruited by a flyer in a local supermarket. The remaining adult participants were acquaintances of SWOV staff interested in participating.

A total of 63 bicyclists participated of whom, 29 were adolescents (M=12.1, SD = .4, Female=17) and 34 adults (M=32.3, SD = 13.3, Female=24). Adolescent cyclists fell in the age range 12 and 13 years old.

To minimize travel demands on the different participant groups, the researchers traveled to different locations (The Hague, Amsterdam and Noordwijk). The test was conducted on site.

Design

A mixed-model design (SPF 2.2 design) was used in this study. This mixed-model consists of the independent between group variable of experience or age (adults and adolescents) and the independent within group variable of condition (secondary task or
control conditions). In practice, this meant that participants did the hazard perception tasks alone half the time, and whilst doing a secondary task with another participant the other half of the time.

The primary dependent variables were two measurements of hazard perception, pressing a button at the right moment and looking at the areas of interest (AOI’s). Secondary dependent variables, variables that were measured in addition to the other two dependent variables, were the visual x-coordinates looked at per clip, the visual y-coordinates looked at per clip, the average fixation duration and the total number of key presses during the video (unrelated to previous hit/miss; not related to the AOI). These dependent variables will be described in the analysis section.

**Motivation for dependent variables**

Within this research project the key press during AOI and fixations longer than 200ms into the AOI are the dependent variables. The SEEV model by Wickens et al. (2001) motivates the decision for fixations, with a focus on AOI’s that correspond with top-down motivated visual scanning behavior since salience AOI’s are expected not to result in a representation of top-down scanning. The key press measure was used to give insight into the understanding or appraisal of a situation as defined in step two in the model of Endsley (1995).

**Stimulus material**

For each participant the same 2 practice videos and 22 test videos were used. The clips were shot in real traffic by two different bicyclists (both interns) with a GOPRO HERO3 camera mounted on the front of their respective bike. These interns collected hours of video footage. This material was analyzed to distill hazardous moments: moments containing latent overt or latent covert hazards. From these hazardous moments in the hours of video footage, many shorter clips containing a hazard were made. These latent hazards never developed into acute threats. Out of the total 55 clips 24 were picked to form the practice and the test videos.

Each clip contained at least one latent hazard. Before the experiment the most important hazard in each video had been determined. This hazard would be the only hazard taken along when scoring the participants (i.e. no points are received for scanning lower importance hazards). This latent hazard can be either overt or covert. The clips lasted between 30 seconds and 1 minute. The hazard was never at the direct start of the video to insure it wouldn’t be missed as a participant got acquainted with the sudden new traffic situation, but
the hazard could be at any other moment in the clip. In Appendix A screen capture of the priority hazard of each of the 22 test clips is shown accompanied with a brief description.

Of each of the video’s the area of interest (AOI) was determined. The area of interest consists of the time frame and target area during the video in which it is especially important to be aware of the main latent hazard. Consequently, afterwards it was possible to measure if participants were aware either visually (fixation pattern) or otherwise (key presses) that a latent hazard was occurring.

Two sets of 11 clips, set A and set B, were then made of the 22 test clips. Additionally, two sets were made with the order of the clips reversed. This was done for counterbalancing objectives like controlling for video order and controlling for the influence of the manipulation (the additional task) on the scores.

The clips were played in the software of the eye tracker called Gazepoint. Between each clip there was a short break with a picture explaining what direction would be biked towards in the next video, followed by a dot in the middle to ensure the starting point of each video would be roughly the same horizontal and vertical coordinate.

**Apparatus**

The clips used in this study were made using a GOPRO HERO3 camera (horizontal angle of 140°) mounted to the front of the bicycles of two interns. They filmed during their normal routines and no situations were planned or set up.

During the experiment the participant was seated on a chair behind a desk. On this desk, there was the monitor. The participant was asked to keep his or her head on a chinrest to insure the eyes kept on a relatively stable position in relation to the eye tracker. The participant was seated in front of a 21.5-inch flat screen monitor (aspect radio 16/9) which was placed 60 centimeters from their eyes on a table. The 60 centimeters were the ideal distance for the eye tracker and it was a comfortable distance as to not being too close nor too far. Figure 3 shows the experimental set up.

The two guides leading the experiment were seated at an angle not to be in the field of view of the participant. This was done to control for a possibly distracting factor of having more stimuli in your field of view.

The eye tracking equipment consisted of a small device place under the monitor. This device is called ‘Gazepoint GP3 Eye Tracker’. By running a calibration session within the
Gazepoint software, this device could measure where a participant was looking during the experiment by examining the participant’s eyes. The clips were played within the Gazepoint software. The output the software allows to analyze the fixation points and duration.

Key presses were also measured within the Gazepoint software. A key press, and the length of the press and the time frame in which the key was held down, were all recorded whenever a participant pressed or held a designated key on the keyboard. For this experiment, it was the right mouse button key under the trackpad, since it was located comfortably on the far right side of the keyboard used in this experiment.

![Figure 3](image)

*Figure 3. The experimental set up with the participant in a chin rest, the monitor and eye tracker at a comfortable distance and the keyboard well-placed for key presses.*

**Secondary task**

This research project contained a secondary task that was balanced out between conditions. The secondary task (the manipulation condition) in this experiment is an animal guessing game. This guessing game requires two players. One person (player 1) who asks questions and one person (player 2) who can respond to them with only ‘yes or no’ answers. The goal of player 1 is to discover what animal player 2 his card reads. The more animals are guessed, the better the game was played. A reward was offered to the pair that guessed the most animals to incentivize player 1 to play the game actively and do his best.

The animals were based on a list of commonly known animals and were printed out on cards. This way player 2 only had to grab the next card to get a new animal, ensuring all
participants had similar animals and ensuring that wasn’t required to come up with animals himself.

Player 2, the one who had to guess the animal and come up with questions, was always the participant doing the hazard perception test at the same time. The goal of this secondary task was to create a condition of extra workload during the hazard perception test. The motivation is to see what the effects of distraction or additional workload is on cyclists.

**Procedure**

Recruitment of the adolescent population by the secondary school led to these participants already having signed an informed consent form before the study. This informed consent form was also signed by their parents, considering these participants were younger than 18 years of age. In contrast the adults signed their informed consent form at the start of the experiment. After this the experiment would commence and follow the same pattern for both groups.

The test required either one or two participants to be in the room for different parts of the experiment. During section 1 and 4 there was only one participant in the testing room. During section 2 and 3 both participants were in the room. A clear overview of the different section is presented in figure 4. The experiment took a maximum of one hour for one pair of participants. The two participants forming a pair were always of the same gender, because this best simulates the presence of peers within the adolescent age group.
Figure 4. The different stages or sections of the experiment and the corresponding participant and condition that was applicable, during the testing cycle of one pair.

The participants would arrive and be asked to wait outside the test room until the assistant experimenter would invite him or her to join. For section 1 participant 1 would be invited in first. If this participant is an adult, the informed consent would be signed. After this the participant is asked to read an introduction to the hazard perception test. The introduction describes the upcoming task, encompassing watching videos and pressing the button as one judges a situation to require additional attention to ensure safe passing, i.e. hazardous. Then the participant is asked to place his or her head on the chinrest, followed by a calibration round of the eye tracker. The two practice videos follow. Following this is the first set of 11 clips.

Then section two starts. The second participant is invited in. Both participants get to read an introduction about the secondary task or manipulation. Then the test guides play one round of the animal guessing game (the manipulation) to make sure the working is clear. Afterwards, participant 1 will start the second set of 11 clips whilst guessing animals with the participant 2 giving feedback on the guesses.

After the set of clips is finished, section 3 starts. Now participant 2 will take the seat of participant 1 behind the computer. They get to read the introduction to the hazard perception test as the test guides clean the chinrest to maintain a high hygienic standard. Following this participant 2 is requested to place his or her head in the chinrest, calibration of the eye tracker
follows and the practice round of two clips is started. Meanwhile participant 1 fills out a questionnaire that my colleague de Geus (2016) used for her research project. When both are done with their respective introduction and questionnaire, participant 2 will get to watch the first set of 11 clips whilst guessing animals as participant 1 gives feedback on the guesses.

Section 4 will start when the first set of 11 clips for participant 2 ended. Participant 1 is given compensation for participating in the experiment (10-euro gift card) and asked to leave. Now participant 2 will commence with the second set of 11 clips without a secondary task. After this participant 2 will fill in the questionnaire and receive compensation for participation. The experiment is now done and all raw data of the two participants have been recorded.

After the experiment the equipment was cleaned to insure hygienic starting conditions at the next session. The recorded eye tracking results were stored on both the test computer and on a backup hard drive. In case there was another session quickly following the previous one, all forms and questionnaires would be laid out in preparation.

Analysis

The Gazepoint software recorded the fixation location and the key presses (in milliseconds) of each participant during each clip. To convert these data into scores, two things were defined per clip, as was shortly explained in the stimuli section. Firstly, the time frames in which a high priority latent hazard occurred were defined for each of the 22 test clips. Secondly the visual area in which this high priority latent hazard occurred was defined: the target area. These formed the area of interest (AOI) per clip, which is used to quantify if participants were aware of the main latent hazard. The guideline in choosing the time frame and target area was to pick the main latent hazard for which the clip was picked in the first place and defining the time frame and target area that optimally captured this hazard. This led to the software producing data that shows if someone clicked during the time frame and if someone looked into the target area (area of interest) for 200ms or more. Additionally, the software produced other measures such as: the mean y-coordinate and standard deviation, the mean x-coordinate and standard deviation, average fixation duration and the total amount of key presses.

The dependent variables, key press and visual behavior, are defined as a hit or a miss. A key press is categorized as a hit when the participant pressed the button, during a hazardous event. If the participants fail to do so, it’s categorized as a miss. The looking behavior of the
participants could also be categorized as a hit or miss. It’s counted as a hit if a participant looked into the AOI for 200ms or more when the AOI was present (scanning pattern measure). To calculate the total scores per variable, the hits were awarded one point each and added up. One point could be scored per item. The resulting sum has a lowest possible score of 0 and a highest possible score of 11 (there are 22 items in total split into two conditions) per condition (secondary task or control conditions). Both variables have the same scoring method.

The key presses give an indication of whether the participant perceived the AOI as a potential hazard and thus requires additional attention, or if they looked into the AOI without considering it to require additional attention. This relates to the concept of ‘looked-but-failed-to-see’ (Herslund & Jørgensen, 2003) and is a way to ascertain if ‘looked-but-failed-to-see’ plays a role within hazard perception among cyclists.
Results

The results section describes the results that support the two main aims. The first aim is to compare adolescents (risk group) and adults (control group) on the measure of hazard perception (key press during the AOI and fixations of 200ms or more in the AOI). The second aim is to set a first step in the development of a hazard perception test that meets both the requirements for validity and reliability.

Results of the experiment

Firstly, the results regarding key presses during AOI’s were analyzed to see if there is a significant difference between adults and adolescents. The maximum score was 11 per condition (key was pressed during all AOI’s) and minimum score was 0 (no key presses during AOI’s). Because of the skewed data, a repeated measures ANOVA could not be conducted.

Therefore, the non-parametric equivalent, the Mann-Whitney test was performed to test between-group differences. The Mann-Whitney’s U test was performed to compare the adults and adolescents in both conditions. The number of key presses (hit/miss) during the area of interest (AOI) by adults was significantly higher than the number of presses by adolescents in the no secondary task condition, $U = 711.50$, $z = 3.078$, $p = .002$, $r = .388$. In the secondary task condition this result was not replicated. No difference between adolescent and adults appeared in the secondary task condition, $U = 410.$, $z = -1.178$, $p = .239$, $r = .148$. Figure 5 shows these results in a graph.
Figure 5. The plot showing the score on key presses during the AOI to indicate they thought the situation was dangerous on the y-axis. It shows four different point representing both adult and adolescent groups whilst performing a secondary task and whilst not comparing a secondary task. Error bars represent +/-1 SE.

Secondly the scores regarding looking in the AOI’s were analyzed by conducting an independent sample t-test and a repeated measures ANOVA. The maximum score was 11 (all AOI’s have been looked at) and minimum score was 0 (no AOI’s have been looked at). For the repeated measures ANOVA the within-subjects factor was the number of AOI’s that were looked at. The between subjects factors were group (adult and adolescent) and condition (secondary task or no secondary task).

The repeated measure ANOVA found no main effect for group, $F(1,47) = 0.218, p = 0.643, \eta^2 = .005$. The interaction effect was not significant, $F(1, 47) = 1.516, p = .224, \eta^2 =$
.031. This is visualized in Figure 6.

*Figure 6.* The plot showing the number of AOI that were fixated on for 200ms on the y-axis. It shows four different point representing both adult and adolescent groups whilst performing a secondary task (distraction) and whilst not performing a secondary task (control). Error bars represent +/1 SE.

**Additional results**

In addition to the measures used for the primary research objectives, more dependent variables had been measured during the experiment: the mean x-coordinate and standard deviation, the mean y-coordinate and standard deviation, average fixation duration and the total amount of key presses. On all these additional measures a repeated measures ANOVA was conducted with the between subjects’ factors being group (adolescent or adult) and condition (with or without a secondary task). The motivation of these test is explorative in nature. The chance exists some of these measures might turn out to be useful measures for future research. Only the between subject factor of group will be reported.

Firstly, the standard deviation on the x-axis during the clips was examined. This measure gives insight into how much the participants scanned the screen from left to right. A low value can be interpreted as staring behavior, whilst a high score on x-axis variation is arguably an active scanner of the periphery.
To examine the visual scanning behavior on the x-axis a mixed between-within subjects’ analysis of variance was conducted. Levene’s test of equality of variance was violated and consequently the scores were transformed logarithmically. One adolescent participant was removed from the analysis as to being an outlier. There was a significant main effect of group, \( F(1,45) = 5.895, p = .019, \eta^2 = .286 \). Adults scanned the x-axis less thoroughly (\( M = .14, SD = .02 \) in control condition; \( M = .12, SD = .02 \) in secondary task condition) than adolescents (\( M = .14, SD = .02 \) in control condition; \( M = .15, SD = .04 \) in secondary task condition).

Secondly the mean standard deviation on the y-axis during the clips was examined. This measure gives insight into how much the participants scanned the screen from up to down. Considering analyzing traffic requires no looking up and down, the measure is arguably irrelevant regarding hazard perception.

To examine the visual scanning behavior on the y-axis a mixed between-within subjects’ analysis of variance was conducted. Levene’s test of equality of variance was violated and consequently the scores were transformed logarithmically. One adolescent participant was removed from the analysis as to being an outlier. There was no a significant main effect of group, \( F(1,46) = 1.215, p = .276, \eta^2 = .026 \).

Thirdly fixation duration was examined. A mixed between-within subjects’ analysis of variance was conducted. Levene’s test of equality of variance was violated and consequently the scores were transformed logarithmically. Two participants were removed from the analysis as to being an outlier. There was no a significant main effect of group, \( F(1,44) = .332, p = .573, \eta^2 = .007 \).

As fourth and last a repeated measures ANOVA was run on the measure ‘total number of key presses’. Levene’s test of equality of variance was violated and consequently the scores were transformed logarithmically. There was a significant main effect of group, \( F(1,61) = 13.85, p = .001, \eta^2 = .185 \). Adults pressed the key more often (\( M = 28.59 \) presses, \( SD = 16.73 \) in control condition; \( M = 27.47 \) presses, \( SD = 10.57 \) in secondary task condition) than adolescents (\( M = 21.38 \) presses, \( SD = 11.09 \) in control condition; \( M = 16.41 \) presses, \( SD = 7.23 \) in secondary task condition).
Results surrounding test development: Key press

A reliability analysis was conducted to calculate the Cronbach’s alpha. The internal consistency of all (22) items regarding key presses during the AOI was $\alpha = .38$. This does not reach Kline’s criterion ($\alpha > .7$) for ability tests (Kline, 1999 as described in Vlakveld, 2014).

In order to assess the discriminant validity of the test a binary logistic regression was conducted with the group as dependent variable and the final scores on key presses during the AOI as the only predictor. The score was a significant predictor ($p < .01$) and the Omnibus test was significant, $\chi^2 = 10.529, p = .001$, with a Nagelkerke $R^2$ of .21. The overall classification accuracy was 60.3% (76.5% for adults and 41.4% for adolescents).

Table 1 contains the percentage of participants that clicked to indicate a feeling of danger or extra attention during an area of interest during an item. The resulting odds ratios are also reported for each item. For instance, an odds ratio of 5,5 indicates that it was 5.5 times more likely that an adult participant pressed a key during the AOI than an adolescent. To test if the scores differed significantly between the adult and the adolescent group, the ‘Fisher’s Exact Test’ was used.

Table 1

The percentage of participants that pressed their key during the AOI per clip for adults and adolescents respectively. Also, it displays the results of the Fisher’s Exact Test and the odds ratio per clip.

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Adult (%)</th>
<th>Adolescent (%)</th>
<th>P (Fisher’s Exact Test)</th>
<th>Odds ratio (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>.848</td>
</tr>
<tr>
<td>A 2</td>
<td>9</td>
<td>7</td>
<td>1</td>
<td>1.306</td>
</tr>
<tr>
<td>A 3</td>
<td>38</td>
<td>17</td>
<td>.094</td>
<td>2.971</td>
</tr>
<tr>
<td>A 4</td>
<td>6</td>
<td>7</td>
<td>.233</td>
<td>.300</td>
</tr>
<tr>
<td>A 5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>.848</td>
</tr>
<tr>
<td>A 6</td>
<td>41</td>
<td>17</td>
<td>.055</td>
<td>3.360</td>
</tr>
<tr>
<td>A 7</td>
<td>26</td>
<td>17</td>
<td>.545</td>
<td>1.728</td>
</tr>
<tr>
<td>A 8</td>
<td>41</td>
<td>14</td>
<td>.025</td>
<td>4.375</td>
</tr>
<tr>
<td>A 9</td>
<td>35</td>
<td>28</td>
<td>.593</td>
<td>1.432</td>
</tr>
<tr>
<td>A 10</td>
<td>6</td>
<td>0</td>
<td>.495</td>
<td>&gt;100</td>
</tr>
<tr>
<td>A 11</td>
<td>24</td>
<td>24</td>
<td>1</td>
<td>.967</td>
</tr>
<tr>
<td>B 1</td>
<td>21</td>
<td>38</td>
<td>1</td>
<td>1.156</td>
</tr>
<tr>
<td>B 2</td>
<td>79</td>
<td>89</td>
<td>1</td>
<td>&gt;100</td>
</tr>
<tr>
<td>B 3</td>
<td>21</td>
<td>17</td>
<td>.678</td>
<td>1.800</td>
</tr>
<tr>
<td>B 4</td>
<td>93</td>
<td>94</td>
<td>.092</td>
<td>4.154</td>
</tr>
<tr>
<td>B 5</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Based on table 1 a top 10 of the items with the highest odds ratio was made. The goal was to analyze if a test consisting of only the 10 strongest items would result in a test that fulfilled the discriminant validity and internal consistency requirements. The top 10 excludes the very high values caused by one party scoring 0%. This top 10 includes A3, A6, A7, A8, A9, B3, B4, B8, B9 and B10. On these items the discriminant validity test and internal consistency tests were run.

A reliability analysis was conducted to calculate the Cronbach’s alpha. The internal consistency of the best 10 items regarding key presses during the AOI was $\alpha = .373$. By deleting 4 items the Cronbach’s alpha could be increased to $\alpha = .442$. This does not reach Kline’s criterion ($\alpha > .7$) for ability tests (Kline, 1999 as described in Vlakveld, 2014).

In order to assess the discriminant validity of the shorter test consisting of the 10 items a binary logistic regression was conducted with the group as dependent variable and the final scores on key presses during the AOI as the only predictor. The score was a significant predictor ($p < .001$) and the omnibus test was significant, $\chi^2 = 18.944$, $p < .001$, with a Nagelkerke $R^2$ of .347. The overall classification accuracy was 68.3% (79.4% for adults and 55.2% for adolescents).

**Results surrounding test development: Fixations**

A reliability analysis was conducted to calculate the Cronbach’s alpha. The internal consistency of all items regarding scanning in the AOI was $\alpha = .11$. This does not reach Kline’s criterion ($\alpha > .7$) for ability tests (Kline, 1999 as described in Vlakveld, 2014).

In order to assess the discriminant validity of the test a binary logistic regression was conducted with the group as dependent variable and the final scores on scanning patterns regarding the AOI as the only predictor. The score was not significant as a predictor ($p = .216$) and the omnibus test wasn’t significant, $\chi^2 = 1.573$, $p = .210$, with a Nagelkerke $R^2$ of .03. The overall classification accuracy was 63.5% (79.4% for adults and 44.8% for adolescents).
Table 2 contains the percentage of participants that looked into the area of interest during an item. The resulting odds ratios are also reported for each item. For instance, an odds ratio of 3 indicates that it was 3 times more likely that an adult participant looked into the AOI than an adolescent. To test if the scores differed significantly between the adult and the adolescent group, the ‘Fisher’s Exact Test’ was used.

Table 2

The percentage of participants that looked in the AOI per clip for adults and adolescents respectively. Additionally, it displays the results of the Fisher’s Exact Test and the odds ratio per clip.

<table>
<thead>
<tr>
<th>Item ID</th>
<th>Adult (%)</th>
<th>Adolescent (%)</th>
<th>P (Fisher’s Exact Test)</th>
<th>Odds ratio (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 1</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>1.571</td>
</tr>
<tr>
<td>A 2</td>
<td>36</td>
<td>17</td>
<td>.210</td>
<td>2.639</td>
</tr>
<tr>
<td>A 3</td>
<td>59</td>
<td>67</td>
<td>.768</td>
<td>.708</td>
</tr>
<tr>
<td>A 4</td>
<td>27</td>
<td>16</td>
<td>.514</td>
<td>1.909</td>
</tr>
<tr>
<td>A 5</td>
<td>42</td>
<td>17</td>
<td>.104</td>
<td>3.571</td>
</tr>
<tr>
<td>A 6</td>
<td>21</td>
<td>20</td>
<td>1</td>
<td>1.091</td>
</tr>
<tr>
<td>A 7</td>
<td>7</td>
<td>14</td>
<td>.638</td>
<td>.444</td>
</tr>
<tr>
<td>A 8</td>
<td>81</td>
<td>83</td>
<td>1</td>
<td>.833</td>
</tr>
<tr>
<td>A 9</td>
<td>4</td>
<td>32</td>
<td>.015</td>
<td>.083</td>
</tr>
<tr>
<td>A 10</td>
<td>15</td>
<td>32</td>
<td>.277</td>
<td>.377</td>
</tr>
<tr>
<td>A 11</td>
<td>24</td>
<td>22</td>
<td>1</td>
<td>1.114</td>
</tr>
<tr>
<td>B 1</td>
<td>21</td>
<td>38</td>
<td>.222</td>
<td>.443</td>
</tr>
<tr>
<td>B 2</td>
<td>79</td>
<td>89</td>
<td>.445</td>
<td>.431</td>
</tr>
<tr>
<td>B 3</td>
<td>21</td>
<td>17</td>
<td>1</td>
<td>1.304</td>
</tr>
<tr>
<td>B 4</td>
<td>93</td>
<td>94</td>
<td>1</td>
<td>.765</td>
</tr>
<tr>
<td>B 5</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>.840</td>
</tr>
<tr>
<td>B 6</td>
<td>89</td>
<td>63</td>
<td>.045</td>
<td>5</td>
</tr>
<tr>
<td>B 7</td>
<td>23</td>
<td>13</td>
<td>.472</td>
<td>2</td>
</tr>
<tr>
<td>B 8</td>
<td>50</td>
<td>50</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B 9</td>
<td>57</td>
<td>56</td>
<td>1</td>
<td>1.048</td>
</tr>
<tr>
<td>B 10</td>
<td>21</td>
<td>22</td>
<td>1</td>
<td>.921</td>
</tr>
<tr>
<td>B 11</td>
<td>11</td>
<td>15</td>
<td>.683</td>
<td>.680</td>
</tr>
</tbody>
</table>

Based on table 2 a top 10 of the items with the highest odds ratio was made. This top 10 includes A1, A2, A4, A5, A6, A11, B3, B6, B7 and B9. On these items the discriminant validity test and internal consistency tests were run.

A reliability analysis was conducted to calculate the Cronbach’s alpha. The internal consistency of the best 10 items regarding scanning in the AOI was \( \alpha = .38 \). By deleting 4
items the Cronbach’s alpha could be increased to $\alpha = .47$. This does not reach Kline’s criterion ($\alpha > .7$) for ability tests (Kline, 1999 as described in Vlakveld, 2014).

In order to assess the discriminant validity of the shorter test consisting of the best 10 items a binary logistic regression was conducted with the group as dependent variable and the final scores on scanning patterns regarding AOI as the only predictor. The score was not significant as a predictor ($p = .052$), but the omnibus test was significant, $\chi^2 = 4.449, p < .05$, with a Nagelkerke $R^2$ of .22. The overall classification accuracy was 72% (69.2% for adults and 75% for adolescents).
Discussion
In the discussion both the results and the research questions will be discussed. The questions were:

1. Do differences exist between high accident risk cyclists and low accident risk cyclists on the higher cognitive ability of hazard perception? What are the effects on experience on hazard perception? If differences do occur, do they show the expected direction?

2. Does the information procession theory explain the relationship between hazard perception scores and accident risk among bicyclists, controlling for the effects of distraction?

3. Which effect does distraction have on hazard perception? Is the effect the same for different risk groups?

4. Does the test manage to distinguish the different experience groups better than chance? Which (kind of) stimuli show to be the most promising in distinguishing low and high risk cyclists? Is the test reliable in the form of internal consistency?

The experiment
The experiment was conducted to give insight into the first question about the differences between experience groups. The results show adults (experienced) outperformed adolescents (low experience) on the key press hazard perception score, but not on the hazard perception score regarding fixations. Consequently the answers are, based on this research project, that differences in perception of hazards between low and high experience groups do not exist, but that differences do exist in indicating if a situation is hazardous. The experienced group is the group shower a higher hazard perception score on this measure. The next paragraphs will go deeper into the relation of these results with the theory.

For the first result about differences in hazard perception, in the form of fixations, the expectation was that adolescent would show a lower hazard perception score. Chinn, Alliot, Sentinella and Williams (2004) report the number of serious injuries and deaths in traffic peaks in the first few years of secondary school, which could be attributed to inexperience and the less developed hazard perception skill, resulting in not perceiving the hazards or perceiving less of the hazards. Additionally, based on the SEEV model by Werneke and Vollrath (2012), the expectation was that adolescents would have a less well developed top-
down visual attention allocation system, resulting in a poorer hazard perception score in the form of missing more (especially covert) hazards. A repeated measure ANOVA comparing the eye tracking hazard perception score between both groups showed no significant effect of group. Analyzing the significant discriminatory items, it is even shown that adolescents outperform adults on item A9, which is a covert hazard (as can be read in Appendix A). These results imply that hazard perception in the form of scanning patterns is not impaired in the adolescent age group. This is in line with SWOV (2013a) showing that teenagers show no clear increase of risk per distance traveled on bike compared to adults. The reasons as to why these results were not in line with expectations and theories might be found in the application of the theories the expectations were based on, and in the role that ‘seeing’ plays within the entire hazard perception process. The explanation could be found in the fact that both groups already had years of experience cycling and stopped showing a significant improvement, in scanning patterns, when training was not involved. This would explain why the results of the scanning behavior are quite similar, but not why there is a peak in accidents at the start of secondary school (Chinn, Alliot, Sentinella & Williams, 2004). Consequently another explanation, as is shown by the model by Endsley (1995), reads that perception is the first step and appraisal the second. Consequently, it is possible that both adolescents and adults perceive the hazards, as is shown in the results of this experiment, but that they appraise them differently. That brings up the second question regarding indicating a situation is hazardous and the question if ‘looked-but-failed-to-see’ occurs within this research.

The second result surrounds the key press measure of hazard perception and whether adults outperform adolescents on indicating there is a hazardous situation present. Based on Jeannot, Kelly and Thompson (2003) their model, the expectation is that, due to their experience, adults are better at step 2: integrating the information with relevant internal knowledge to create a mental picture of the current situation. This results in a better judgement whether or not a situation is hazardous. Due to the skewed data, the Mann-Whitney’s $U$ test was conducted to compare the scores of adults and adolescents in both secondary task conditions (control and secondary task). The results show that the during the secondary task condition the scores did not significantly differ, in contrast to the significant difference in the control condition. These results imply that adults outperform adolescents when not actively distracted, but also that the scores diverge closer in a secondary task condition. Considering this paper does not focus on the influence of distraction, the first result in the control condition will be discussed. This finding implies that adults have a more
developed hazard perception skill in the control condition, which is in line with the expectations based on theory. As put forward by Arnett (1994), adolescents show differences from adults on the area of emotion and appraisal. Arnett (1994) showed that adolescents are more sensation seeking in traffic. In line with this trend, it is also a possibility they do see the risk, but they are more risk accepting. This would be in line with the ‘looked-but-failed-to-see’ phenomenon (Herslund & Jørgensen, 2003). Another possibility is that adolescents simply do not see the risk. Either way, this experiment shows that adolescents are poorer than adults in appraising the hazardous situations as hazardous, especially considering the fixation data shows no significant difference at step 1 the model by Jeannot, Kelly and Thompson (2003), regarding perception.

**Test development**

The secondary goal within this research project was to set the first steps towards developing a psychometrically valid hazard perception test for bicyclists that distinguishes between experienced and not experienced cyclists. The fourth researched question has been raised in order to assess this subject. These are the questions as formulated at the fourth question in the research questions section:

- Does the test manage to distinguish the different experience groups better than chance?
- Which (kind of) stimuli show to be the most promising in distinguishing low and high risk cyclists?
- Is the test reliable in the form of internal consistency?

To test the internal consistency question, reliability analyses were conducted to analyze if Cronbach’s alpha would be equal to, or higher than, .7 for both different test measures (fixations and key presses) and sizes (all 22 items and 10 strongest items). Only one test is required at the end, so the analysis of multiple formats is used to point to the preferred version for future research. The expectation was that all of these tests are reliable. However, the analyses showed that none of the tests reached the criterion that points to a reliable test. Due to the explorative nature of this research, it’s neither in line nor not in line with previous research into hazard perception tests for bicyclists. Compared to the hazard perception tests for automotive vehicle operators, however, one can assume that psychometric validity is a distinct possibility for these types of tests and that this test has failed to achieve the required reliability. This result indicates that there is something about this experiment that makes it
weak. This requires explanation, which can be explained by analyzing the items and what causes them not to be internally consistent. The most likely culprit is the stimulus material. As can be seen in table 1 and table 2, there are items that prove to be significant distinguishers between groups in both directions, in contrast to many items that do not significantly distinguish between groups. Besides that, the reliability also suffered from the direction of the results. Items showed results in both directions: adults outperformed adolescents in some, whilst other items showed a reverse trend. Using Appendix A it is possible to trace back the overt and covert items, but both these categories show results pointing in both directions and the overt category suffers from a rather low N. Consequently, the role that covert and overt hazards are expected to play based on the SEEV model (Werneke & Vollrath, 2012) does not return in these results in the form of a reliable test. Explaining these different results remains in the realm of guessing, due to the lack of a theoretical framework into cyclist hazard perception. Based on the available frameworks about situational awareness, such as the one by Jeannot, Kelly and Thompson (2003), the most likely explanation is that adolescents and adults do indeed update their mental of their surroundings in different ways, as expected, but the factors that form these mental models are still unknown in the area of cyclists and are not accentuated within the items of this experiment. This brings up the question what makes certain items appropriate for hazard perception tests.

The other two questions focus on what makes items strong and if they are strong enough combined to distinguish between groups. A strong item is one that can discriminate between adults and adolescents much better than chance, chance being 50%. To test whether or not the test contained strong items, the discriminant validity of the four test types on distinguishing between adults and adolescents was calculated before looking at individual items. The two key press measure tests reached significance in discriminant validity, the two scanning pattern measure tests did not. However, it should be noted that the significant scores were not much better than chance, being 63.5% and 68.3% respectively. Since chance is 50%, this seems to imply that the tests might be significant discriminators, but not strong discriminators. Consequently one could argue that the discriminant validity is weak, too weak, to classify these tests as a strong enough discriminator compared to chance. The explanation can likely be found again in table 1 and table 2. Both adults and adolescents scored quite similar on most items. The result is that discrimination is based on only a few strong items, resulting in an overall discriminant validity not much better than chance. This brings us back to the question what makes these few individual items so strong compared to the rest.
The strong, discriminatory, items distinguish are identified by looking which items are significant discriminators in Table 1 and Table 2. There are 4 significant items. For fixations, these are clips A9 and B6. For key presses, these are clips A8 and B8. Considering this is a small pool of data from which to draw conclusions, it is hard to identify the qualities that make these items strong. A detectable trend is, nonetheless, that adults are more likely to press the key indicating a hazardous situation when there is a covert hazard, as is seen in A8 and B8. In contrast to expectations based on the SEEV model (Werneke & Vollrath, 2012), A9 and B6 show that adolescents are more likely to scan a covert location from which an individual already appeared and that adults are more likely to perceive overt bicyclists with low salience, but who will cross your path in the future. This could point to adults being better at picking out the most important overt hazard in a crowd of traffic. Additionally, the other item could point to focus on covert locations of adolescents being triggered by other users appearing from that location, more than it does for adults. Either way, strong items are characterized by demanding a degree of top-down understanding of what can be expected next in traffic situations with covert hazards, or a top-down understanding of where attention should focus on in busy situations with overt hazards.

**Conclusion**

The main goal of this study was to research differences in hazard perception between adolescent and adult bicyclists. The expectation was that adults, due to their cycling experience, would perform better on hazard perception than adolescents, a low cyclist and high accident risk group. The conclusion is that the adolescent (risk) group does score lower on hazard perception than adult bicyclists on a key press measure whilst not being distracted. On a fixation or eye tracking measure, in contrast, adults and adolescents do not show any significant differences on their hazard perception scores. In conclusion adolescents and adults perceive (see) the same amount of hazards, but adults significantly more appraise the situations as hazardous.

The secondary goal was to set a first step towards developing a cyclist hazard perception test. To achieve this, the goal was to develop, or set the first steps toward, a psychometrically valid test which can measure hazard perception in bicyclists and additionally distinguish adults (experienced) from adolescent (less experienced) cyclists. The test(s) used in this study were not psychometrically valid, considering their reliability was insufficient, even when only taking a smaller sample of the strongest items. The items in this study were too weak to form a psychometrically valid test.
Additionally a test would be strong when it successfully distinguishes adult from adolescent cyclists. This study show promising significant, but weak, results. The test with the measure key press could distinguish adults from adolescents only slightly better than chance. The other measure, scanning patterns regarding the AOI, did not replicate this result and not succeed in significantly distinguishing between the groups.

This study shows that a key press measure is preferred over a fixation measure when distinguish between the novice and experienced bicyclist groups. Additionally, it shows that adolescents and adults do differ on hazard perception on the key press measure. Nonetheless, this study also shows that this test isn’t reliable and requires improvement.
Limitations and recommendations

Limitations

This study brings to light multiple limitations. The most important one is that sitting in a research set up is not bicycling. It does not replicate a real situation, it only aspired to come close to it. Effects on doing physical exercise could significantly affect the measures of this test.

Another limitation relating to the reality of the test compared to cycling, is that the experimental setting didn’t offer a wide visual environment that cyclists normally experience, but rather had only one screen as visual input. Due to the horizontal 140° angle of the video material, participants couldn’t look to their direct right and left, resulting in altered visual behavior.

Another limitation is that there is no framework yet for deciding on the stimulus material. Since there is no clear guideline, nor much research on this subject yet, distinguishing qualitatively high material from lower quality material maintains a subjective task.

A limiting factor surrounding eye tracking is the fact that eye tracking suffers from a lot of missing data. Consequently, researchers have to manually make sure that the eye tracking data is correct on the moments that matter and otherwise filter the participants out of that measurement. Since a higher n is always preferred, eye tracking equipment was limiting factor in this study.

One participant (an adolescent) didn’t show up. Because a part of the experiment requires another player of the same gender, the test leader (age 22) of the same gender functioned as the second participant. Because the role of this second player is only to give yes or no answers to questions, and the test leader was also of the same gender, the assumption was made that the adolescent participant would have scored in a comparable fashion to a same aged second player and was consequently taken along in the analysis.

Recommendations

Based on this research, the principal recommendation for future research is to focus on the key press measure compared to eye tracking measures. Because the eye tracking measure showed different results that were not due to differences in experience, it isn’t an interesting measure in differing between groups.

On stimulus material, the recommendation is to select new stimulus material based on a clear theoretical framework. Because the method used in this study, using the view of experts combined with the opinion of a panel, did result in quite unexpected and diverse
results in item strength. By taking a more standardized approach, future research can learn what makes stimulus material strong or weak.

Another recommendation is to consider changing the position from which material is recorded, and the method in which it is presented. Because the current ‘film’ set up taken from the bicycle, participants reported having trouble looking over bushes they could normally look over. They also reported wanting to look at the left or right by moving their heads, which has no use in the current setting with a monitor. Future research should improve on this by filming from a higher viewpoint and by considering offering a wide-angle method of showing stimulus material.

Future research must find low to no experience groups regarding cycling. Since Dutch adolescents generally already have years of experience cycling, it might not be the best representation of the low experience group. Comparisons using populations with no experience will give more insight into the development patterns of eye scanning behavior.
References


Centraal Bureau voor de Statistiek. (2016a, 2 juli). *Gemiddelde bevolking; geslacht, leeftijd, burgerlijke staat en regio* [Dataset]. Geraadpleegd van http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=70233NED&D1=0&D2=0,13-25&D3=0&D4=1&HDR=G3&STB=G1,G2,T&VW=T

Centraal Bureau voor de Statistiek. (2016b, 2 juli). *Overledenen; doden door verkeersongeval in Nederland, wijze van deelname* [Dataset]. Geraadpleegd van http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=71936ned&D1=0,2,6&D2=0&D3=a&D4=9-19&HDR=T&STB=G1,G2,G3&VW=T


McKenna, F.P., & Farrand, P. (1999). The role of automaticity in driving, in G.B.Grayson (Ed.,), Behavioural Research in Road Safety IX. Transport Research Laboratory, Crowthorne, UK.


Appendix A: Short descriptions of the stimulus material by M.A. Knoll, 2016

Screen captures and brief descriptions of the clips used in the experiment (excluding the practice clips). Each screen capture relates to the hazard that was scored.

Clip 1 (set A; clip 1) (Covert)

The view on possible pedestrians working around the blinking parked van is blocked by the van. Possible pedestrians might jump in from the front of the van. When cycling past the van there stands a person just in front of it. The latent hazard consists of the van blocking your sight as you pass it and the pedestrian that stands in front of it.

Clip 2 (set A; clip 2) (Covert)

The street to the right already had multiple cyclists coming out of it. More bicyclists could possibly come from the street on the right as you get closer to the street. The latent hazard consists of possible cyclists or cars coming from the right.

Clip 3 (set A; clip 3) (Covert)

The bus on the right is blocking the sight on possible traffic behind it. The latent hazard consists of possible traffic appearing from behind the bus.
On the sidewalk to the right a playing child is running around. The child might run into the bike path without taking into account a bicyclist like you. The latent hazard is the playing child.

A good view on potential traffic from the street to the right is blocked by the building. In the video a bicyclist will come out and nearly cut you off. The latent hazard in this video is the possible traffic that might be hidden by the building.

The man to the right walks towards the road in a high pace. If he continues to do so he might be on a possible collision course with you. The latent hazard in this video is this man.
Clip 7 (set A; clip 7) (Covert)

The keen observer will notice pedestrians walking from the left to the right becoming hidden by the van. As you will pass the van, these pedestrians will be on your collision course. The latent hazard are these pedestrians hidden by the van.

Clip 8 (set A; clip 8) (Covert)

The truck is parked on your bike path to unload goods. It is impossible to see if traffic is coming from behind the truck. The latent hazard is the possible traffic hidden by the truck.

Clip 9 (set A; clip 9) (Covert)

A man already appeared from the opening in the scaffolding. Another person might be hidden by the wall and come from that opening as well. The latent hazard in this clip are possible other pedestrians that can come from the right.
The parking garage in the busy shopping street to the right already had a car going in it. Cars can also get out.
The latent hazard in this clip are the possible cars that might come out of the parking garage.

In this clip you enter a busy crossing. Traffic can come from any direction. The most important latent hazards
are traffic from the right (obscured by the building) and traffic from the left (obscured by the truck). The latent
hazard in this video are potential pedestrians from the right, since they have right of way and are rather close
to your bike route.

In this clip you pass a van standing still on the sidewalk with blinking lights. The van could block your sight on
pedestrians or workers around the van. The latent hazard are potential pedestrians around the van.
Clip 13 (set B; clip 2) (Covert)

A car already came out of the street to the right. Your sight on this street is blocked by plants. The latent hazard in this clip is the potential other traffic that might come from the right.

Clip 14 (set B; clip 3) (Covert)

In this clip you enter a busy crossing with traffic coming from three side streets. Although the left one is more salient due to a car, it is important to also make sure there isn’t traffic coming from the right. The view on the right is blocked as you pass a truck standing still. The latent hazard here is possible traffic from the two side roads on the right.

Clip 15 (set B; clip 4) (Overt)

In this clip there are three bicyclists coming from the other side. They are cycling next to each other, making it hard for you to pass. The latent hazard in this video are the oncoming cyclists.
In this clip the view to the right is blocked by plants. There can come traffic from this side. The latent hazard in this video is possible traffic coming from the right.

In this clip multiple cyclists are coming from the right. They are on a collision course with you and require attention. The latent hazard in this video are the cyclists to the right.

In this video there are cyclists coming from the left. They disappear behind the truck blocking your sight. Nonetheless they are still on a possible collision course with you. The latent hazard in this video are the cyclists coming from the left.
Clip 19 (set B; clip 8) (Covert)

The van on the right is parked to unload goods. There already were some pedestrians passing it via the right earlier. They might come out in front of the van as you pass it. The latent hazard in this clip are the pedestrians hidden from sight behind the van.

Clip 20 (set B; clip 9) (Covert)

The van on the right has stopped and has people getting in and out of the van. There might be people coming out as you pass it and people might pass it at the front. The latent hazard here are the hidden potential pedestrians around the van.

Clip 21 (set B; clip 10) (Overt)

A busy crossing with traffic from both the right and left is being approached. All of these road users require attention to safely pass this traffic point. The latent hazard in this video is the car from the right.
In this clip you near a crossing. The cars to your left are standing still, suggesting traffic from other sides might have the green light. This means there can be bikes and scooters coming from the left (in front of the cars). The cars block the view of potential traffic from the left. The latent hazard is the potential traffic coming from the left.