Universiteit Leiden

Psychologie
Faculteit der Sociale Wetenschappen

Explaining risk-taking behavior of vehicle drivers: Wilde's risk homeostasis theory and the role of self-efficacy in a video game.

Name: Gosse Mol

Supervisor: Dr. Jop Groeneweg

Second reader:

Cognitive Psychology

Thesis Msci Applied Cognitive Psychology

Introduction

To take risk is to take part in some activity whereof the outcome is uncertain. Life itself is a risky endeavor and taking risk is an inevitable part of life. From an evolutionary standpoint, taking the right amount of risk is an essential behavior. Often, risky activities have more beneficial outcomes than safe activities. Back when hunting and gathering were the foremost methods of collecting food, the individuals that would hunt the largest animals would gain the esteem of the group. While hunting large animals is more risky than hunting smaller ones, it will also provide the group with more food. Individuals that are in good esteem of the group would also be more likely to attract the most and fittest partners, which increases their chances to successfully further their genes.

Risk is never completely absent in any activity. Even something as simple and safe as stepping out of your bed poses some sort of risk in the form of getting caught in your bedsheets and tripping. Removing all risk from any activity is unrealistic. In that case, the proposed solution for reducing tripping by bedsheet accidents would be to forbid leaving the bed. People take the risk of getting out of bed because they believe themselves capable of coping with the situation and have feelings of self-efficacy (Bandura, 1997). It should not be the goal of safety regulating entities to remove risk. The discussion should rather be about optimizing the amount of risk. Sadly, applying this reasoning in safety regulations is problematic at best from an ethical standpoint. It may prove hard to defend in court and public discussions why some activity should not be forbidden, while it is obviously risky. While more factors than discussed above are involved, a good example is discussions on the legalization of drugs.

Humans optimize the risk they are willing to take through some internal weighing of costs and benefits. For example, when driving, the main goal is to get from one location to another. A secondary goal is to get there in good time. Traveling over a few kilometers, especially in a car, should take no longer than an hour at most. Almost anyone is willing to get into a car and drive at some speed. Getting there slowly is not as useful as getting there in good time. Wilde (1982a) proposed that the amount of risk anyone is willing to take depends on the utility of the behavior. This target level of risk (target risk) is determined by four utility factors. Target risk is higher when 1 and 4 are higher and 2 and 3 are lower.

1) Expected benefits of risky behavior alternatives (getting there sooner).
2) Expected costs of risky behavior alternatives (getting yourself injured in an accident).
3) Expected benefits of safe behavior alternatives (getting there cheaper by avoiding fines).
4) Expected costs of safe behavior alternatives (getting there later).

This experiment concerns a theory that describes the above risk optimization process, which will be discussed below. The study experiment described is part of a series of experiments. Skill, as one of the many psychological factors affecting risk optimization, is investigated in this study experiment. A secondary, and related, objective is the investigation of a questionnaire that measures self-efficacy in video games, which was put together for this study.

The theory of risk homeostasis

People infer the riskiness of any situation or activity. This inference is often an imperfect judgement of the factual risk. A range of cognitive and perceptual factors may disturb the judgement quality of humans. For an introduction on this topic, read Kahneman and Tversky (2011). People compare the perceived risk of any situation or activity to their target risk and choose or adjust their behavior accordingly. In situations or activities that are safe, people are inclined to compensate by taking more risk. Risky behavior often has increased expected benefits. This is also true for the other way around. Drivers slow down when there is a bend in the road or when the road suddenly becomes narrower, even if it is factually unnecessary to do so (Adams, 1988). There is empirical evidence for risk compensation behavior in traffic accident data as well. Some changes in legislation or safety standards did not reduce the number of accidents because drivers were willing to take more risk (Wilde, 1982a).

Wilde (1982a) conceptualizes the aforementioned behavior in a task analysis model of driver behavior. Drivers compare the perceived risk of the situation or activity with their internal target level of risk at a high frequency. This comparison leads to a change in behavior that compensates for factors and elements that reduce or increase risk on the short term. Wilde (1982a) proposed an explanation for this compensation, and states that it is a mechanism in a homeostatic process that works on the long term. The theory of Risk Homeostasis (RHT) is an explanation for empirical findings that although safety regulations in traffic are becoming stricter, the average accident rate over a few years does not decline.

Just like a thermostat adjusting the heating of the water in a central heating system to a level that keeps the room temperature at the set level, the population adjusts its behavior so that the perceived level of risk matches the target level of risk. When the system increases the heat, it is responding to a loss of warmth in the room. This compensation is an adjustment that changes the temperature of the room in short term, but keeps the average temperature across a longer period constant, in order for the temperature to match the target. This is a homeostatic process because temporary deviations from the average will tend to return to the mean through a feedback loop. A thermometer measures the lower room temperature. The thermostat has a threshold value and signals the heater to turn on when the measured temperature is lower.

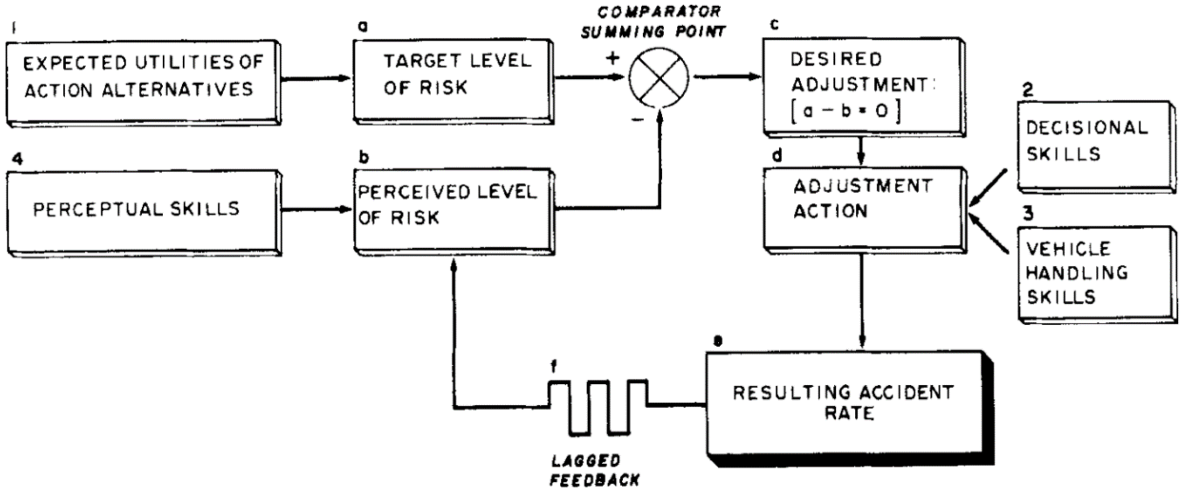

Figure 1. A model of RHT, as proposed by Wilde (1982a).

RHT controversy

RHT has caused much discussion, of which Adams (1988) offers a good overview. The main problem with the theory is that none of the variables in the theory can be accurately measured in the real world. This is true for both authors discussing evidence in favor of RHT as well as authors discussing evidence that refutes RHT. Evans (1986), for example, makes the following claim.

Based on the analysis presented it is concluded that the risk homeostasis theory should be rejected because there is: (1) no convincing evidence supporting it, and (2) much evidence refuting it.

Evans (1986) discusses traffic accident data, but only presents data from fatal accidents, dismissing the fact that this disregards a huge amount of traffic accident data. Although the occurrence of fatal accidents shows a downward trend over time, this is a poor measurement of risk-taking behavior, and can hardly be presented as evidence that refutes RHT.

Other empirical findings show that changes in legislation have short-term effects on traffic accident rates, but the average over the years does not change. Adams (1988) describes several different situations in which objectively safer environments lead to more accidents than subjectively dangerous environments. Firstly, fewer accidents take place in so-called adventure playgrounds compared to safer playgrounds where surfaces are smooth and rounded and equipment is surrounded by cushioning material (Hurtwood, 1968; as described in Adams, 1988). Secondly, the number of fatal accidents and the severity of accidents reduced when road conditions were worse through snow and ice. Thirdly, when Sweden changed to driving on the right side of the road, the number of road accident fatalities were reduced to 40% of the level before the change (Adams, 1985; as described in Adams, 1988).

Despite claims of findings for or against RHT, Adams (1988) argues that RHT is not falsifiable for several reasons. Firstly, because RHT has circular reasoning in the light of evidence that might support it. To quote Adams (1988).

We identify a person with a high target level of risk by his high level of accidents, and we explain his high level of accidents by his high target level of risk.

This is tied to the second point that, in the real world, risk cannot be measured directly. Risky behavior can only be measured by how many accidents have happened or by asking people to report on their own measures of risk-taking. Thirdly, it is impossible to predict the duration of the lagged feedback as depicted in the original model proposed by Wilde (1982a; Figure 1). This is a problem because there may be uncertainty whether enough time has passed before the feedback could have had its effects. Especially on the population scale, feedback could have its effects long after the change or event through a long chain of cause and effect relationships. The feedback could be so lagged that the homeostatic correction could take place several tens of years later (Adams, 1988). Fourthly, it is impossible to find all displacement effects. Displacement is the expression of risky behavior in another activity to compensate for reduced risk-taking in the original activity. Researchers would have to search the entire universe for displacement effects to state with confidence that a homeostatic process is not occurring. Displacement effects were measured when stricter laws on alcohol sale would increase problems from increased ecstasy consumption (Adams, 1988).

RHT in experimental context

Somewhat in line with Adams (1988), Wilde et al. (1985; as described in Jackson & Blackman, 1994) argued that correlational field data is no good to pinpoint any RHT effects. RHT has too many factors that cannot be measured and is not falsifiable in light of correlational data. An experimental setting, on the other hand, could prove to be a better method for confirming the theory. However, there are issues that arise in experimental contexts. One issue that might become more problematic is that of a time constraint of an experiment. Even in real life situations, the lagged feedback could lead to an unknown duration of the feedback loop. On this point, experimental conditions may hardly be any better to measure all feedback effects on the perceived risk of the situation. However, an experiment does offer the possibility to let participants have a timewise collapsed experience. Risky events such as accidents or near misses could be simulated in a shorter period than would occur in real life (Glendon, Hoyes, Haigney & Taylor, 1996).

An experimental setting could prove to be a better method of investigation than real world data on several other counts. For one, it is much easier to measure and control other variables in an experiment. Secondly, experimental research makes it possible to find out the workings of various cognitive and behavioral factors in RHT. Wilde (1982a) proposed a model for driver behavior, which shows the working of and factors affecting risk compensation (Figure 2).

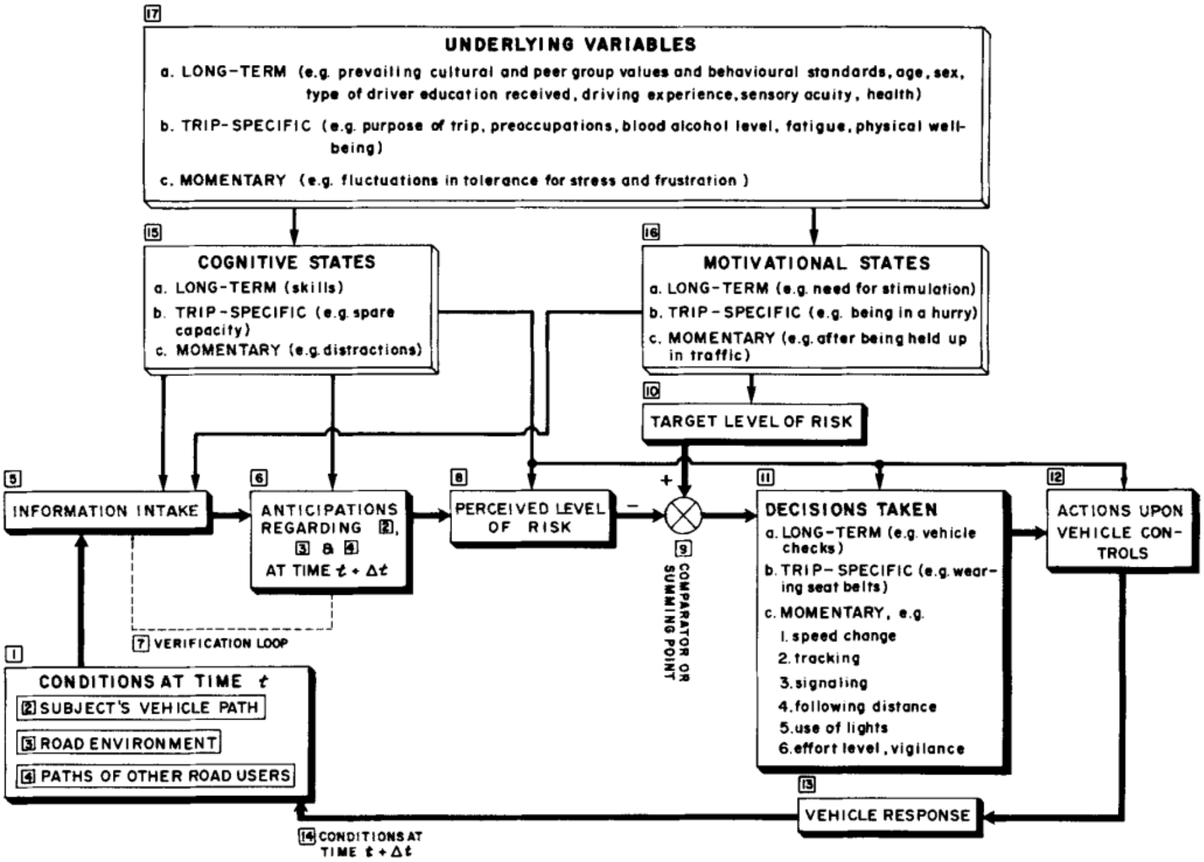

Figure 2. A simplified model for explaining driver behavior (Wilde, 1982a).

Experiments enable the identification of behavioral and situational pathways through which choices are made or how and under what conditions the target level and perceived level of risk are compared. There are some relevant psychological dimensions in RHT that can only be measured in an experimental setting (Glendon et al., 1996). Correlational data from the real world environment and anecdotal examples can hardly be used to draw any of the conclusions that are interesting for the confirmation of the theory. Even though RHT concerns the behavior of the population, in order to make meaningful predictions, it is important to understand what drives the behaviors of individuals.

Skill and self-efficacy

As one of the factors in driver behavior, skill is an important factor in a driver's cognitive state (Figure 2). As discussed in the second paragraph, risk taking is inevitable to achieve anything. Atkinson (1964; as described in Sorrentin, Hewitt, & Raso-Knott, 1992) originally developed a theory of achievement motivation, which describes two types of individuals in relationship to risk-taking. Success-oriented individuals value success greater than failure while failure-threatened individuals value avoiding failure greater than achieving success. Sorrentin, Hewitt, and Raso-Knott (1992) investigate such informational and affective influences on risk-taking behavior in a game. The authors investigate two types of games in a series of studies. The first type is a skilled game, which relies on some ability of the player. The second type is a chance (or unskilled) game, which relies on chance and is nondiagnostic of ability. Most real-world activities that are of interest to this study are skilled activities, such as driving. A participant's type of achievement motivation was found to be predictive of risk-taking in a skilled game (Sorrentin, Hewitt, & Raso-Knott, 1992), while other research into the effects of achievement related motives in chance games have returned negative results. This suggests that achievement related motives, such as wanting to win in a game, are somehow related to whether one can use their skill in a game.

In a series of experiments, Wilde (1988) found that skill and risk-taking were independent. Participants did incur real risk, either financially or in terms of social esteem. However, Wilde (1988) acknowledges that risk-taking might have different workings in other domains. In a study investigating the effects of experience, ability, and self-efficacy on risk-taking, the researchers found suggestions for self-efficacy to play a mediating role. Llewellyn and Sanchez (2008) investigated the effects of self-efficacy and the sensation seeking traits on risk-taking in risky sports, such as rock climbing. More experienced rock climbers had higher levels of self-efficacy, which motivated them to take more risk. Higher self-efficacy scores were positively associated with risk-taking in rock climbing (Llewellyn & Sanchez, 2008). Another study investigated risk taking among male high risk sports, finding that extreme risk takers had higher levels of self-efficacy than high risk takers (Slanger & Rudestam, 1997; as described in Llewellyn & Sanchez, 2008).

Current study

The experiment that is part of this study attempts to measure effects of RHT. The experiment is part of a series of investigations concerning RHT. The risk-taking of participants is measured in a simple videogame, specially created for the experiment. In this study, a newly developed measurement of self-efficacy in videogames is considered (Appendix A). This study aims to find an effect of self-efficacy on risk-taking behavior, and its role in RHT. As a validation method, the predictive value of scores on the questionnaires regarding performance in the videogame will be investigated. Trepte and Reinecke (2011) showed that performance and self-efficacy are related. In their study, self-efficacy in a game was shown to be a mediator between performance in the game and game enjoyment. The mediator, self-efficacy, can therefore be predicted by performance. However, the research method does not exclude that it is in fact the other way around. Self-efficacy scores could be constant, whether it is measured before or after playing the game. There is a well-understood problem with asking participants to rate their own self-efficacy before playing the game. It could lead to unwanted desirability effects on the performance in the game. While the score participants achieve in the game is a good measurement of performance, it cannot be used to find effects of skill on risk-taking behavior, because score is a product of time and speed, of which the latter is a measurement of risk-taking behavior. Speed and score, therefore, are not independent measurements.

In summary, this study experiment measures risk-taking behavior in a video game. It is hypothesized that players are willing to take relatively more risk when their situation in the video game is safe compared to when it is dangerous (H1). The safety of the situation is determined by how many collisions the player could make before losing their ship at any point. To explore how risk-taking behavior might change or develop over time there are several conditions, which differ in respect to the amount of shields that players get at the start of a round. Self-efficacy in videogames is measured with a newly developed self-efficacy scale. In line with the secondary objective, it is hypothesized that self-efficacy in videogames is a reliable scale and positively predicts performance in the game, measured by score (H2). Additionally, it is hypothesized that there is a positive effect of self-efficacy on risk-taking behavior (H3). Finally, it is hypothesized that the risk homeostasis effect is stronger for people with higher self-efficacy (H4).

Method

This experiment used the free version 8.1 of GameMaker (https://www.yoyogames.com) to develop the game for the experiment. This version is no longer available for download from the official site but is still available for download from large file hosting websites, such as Softonic (http://en.softonic.com/s/game-maker-8.1). The game was run on university workstation computers running Windows 7. The computers have 21” displays running at 60Hz refresh rate and set to full HD resolution (1920x1080). The game has a resolution of 800x600 pixels. It is displayed centered on the screen in a fullscreen black window. During execution, the game writes information to a log file containing the amount of shields players have left, the position of the ship, the speed of the ship, the location of the closest meteor, and the location of the first meteor in a direct path of the ship. This information could be used to calculate several measures of risk-taking behavior, such as the speed, how closely players dared to pass meteors, and how closely participants allowed meteors in path to get before dodging them. A list of several questions to assess self-efficacy in games (Appendix A) was developed, taking into account items discussed in the guide for constructing self-efficacy scales by Bandura (2006). The items on the short questionnaire were entered into a Qualtrics (http://www.qualtrics.com/) questionnaire. This questionnaire was displayed and could be answered on the computer by opening the questionnaire link in a Chrome (Google) web-browser window.

Participants (N=181) were randomly selected students at the social science faculty and acquaintances of the researchers. There were no exclusion criteria for participation. When participants agreed on participation, they could make an appointment with the research team. Participants were reminded of participation one day before the experiment through email. The experiment took place in a computer classroom in the faculty of Social Sciences at Leiden University. The classroom had over 20 workstation computers. Up to ten participants could enter a session, effectively allowing one empty computer between every participant to reduce social desirability effects from being able to see each other’s screens. When entering the classroom all participants were welcomed and asked to sign a consent form. Instructions for the experiment and the game are given in a presentation by the research staff. Participants were asked to be quiet and focus on their own screen. Participants received the following instructions for the game.

The game is about a little spaceship in a galaxy not so far away underway to deliver very valuable cargo. If you reach your destination sooner you will get more points. Unfortunately, the ship runs into a thick cloud of meteors. You are the ship's captain and you have to stay on your toes to dodge the danger and get through.

Control instructions were presented in the first screen of the game. When clicking continue, participants would get to play a practice round. Following the practice round, participants would play through five different rounds in a random order. In the game, players navigate a spaceship (90 by 40 pixels) through a thicket of oncoming meteors (each 60 by 60 pixels). The game is a flat two dimensional area where players can control the speed at which meteors fly toward the ship and the vertical position of the ship on screen. The horizontal position of the ship was fixed at 109 pixels from the left side of the screen. A collision with a meteor triggers an animation of an activating shield to protect the ship and removes one shield from the set of shields players receive at the start of a round. When players have lost their shields, another collision triggers an animation of their ship exploding. At every end of a round participants were prompted to click continue to start the next round. A round was completed by losing all shields and destroying the ship or by making it to the end by staying alive for four minutes. At the end of the five rounds, the game closed and the computer switched focus to the questionnaire. Participants were free to collect their reward for participation and leave when they were done with the questionnaire.

The experiment has a double-blind randomized design with five within-subject conditions. The within-subject conditions were different in respect to the amount of shields participants had during the game. Participants had zero, one, three, or five shields, with three shields in two conditions. In one of these two conditions, the actual number of shields is unknown to the participant. These five within-subject conditions translate to seventeen shield situations. Measurements of risk-taking behavior are recorded and averaged over every shield situation, resulting in seventeen within-subject variables per risk-taking measurement variable. For example, the one-shield condition has two shield situations, namely one shield and zero shields left.

Results

181 participants participated in the study. For seven participants something went wrong with the game. A participant that was always flying at the slowest speed was excluded. This means that he or she did not once attempt to fly faster, while the slowest speed is very slow. The goal of the experiment was for participants to match the perceived risk in the game to their internal target level of risk. When participants did not even try a different speed even once, it was assume that they did not understand the instructions. The data concerning the amount of playing time in which participants played the game for each shield there was some noticeable gameplay behavior that warrants additional exclusion from the dataset. Some participants played to finish the game as soon as they could by purposefully flying their ship into meteors. Finally, one participant skipped many items in the questionnaire by filling in nonsense values.

This leaves 168 participants for final analysis, with 47 being male (28.0%) and 121 female (72.0%). Age of participants was hardly normally distributed with a large peak around the mean (M = 22.3, SD = 4.0, Kurtosis = 34.6, Skewness = 4.5) and two outliers in the high end (ages 40 and 57, while all other ages are 30 and lower). Scores in the game were slightly skewed, having a tail towards the higher scores (M = 1938, SD = 1177, Kurtosis = 1.8, Skewness = 1.4). Measurements of risk-taking behavior are the mean speed at which participants dared to pace their ship and the mean time to collision (ttc) participants had to incoming meteors. Lower values of ttc mean that participants moved their ship out of the way of incoming meteors later and this with less margin for error. It should be noted that ttc is a function of speed and distance, thusly related to speed.

Initial exploration of the data shows a problem. The averages of the first shield that participants lose are heavily influenced by the first few seconds of the start of a game session. At the start, the spaceship flies at the lowest speed and with no meteors nearby by default. The spaceship does not change speed when it is hit by a meteor. This causes the averages of the first shields to be severely biased (see Figure 3 and Figure 4). This problem is in fact a major point of discussion from other reports concerning this study. In order to avoid this, the statistics from every first shield that participants played and lost were removed from analysis. This means that the zero-shields condition is dropped from analysis. The four situations in the five-shields condition, two situations in the three and unknown-shields conditions, and one situation in the one-shield condition remain. This leaves twelve repeated measurements for each mean speed and mean ttc.

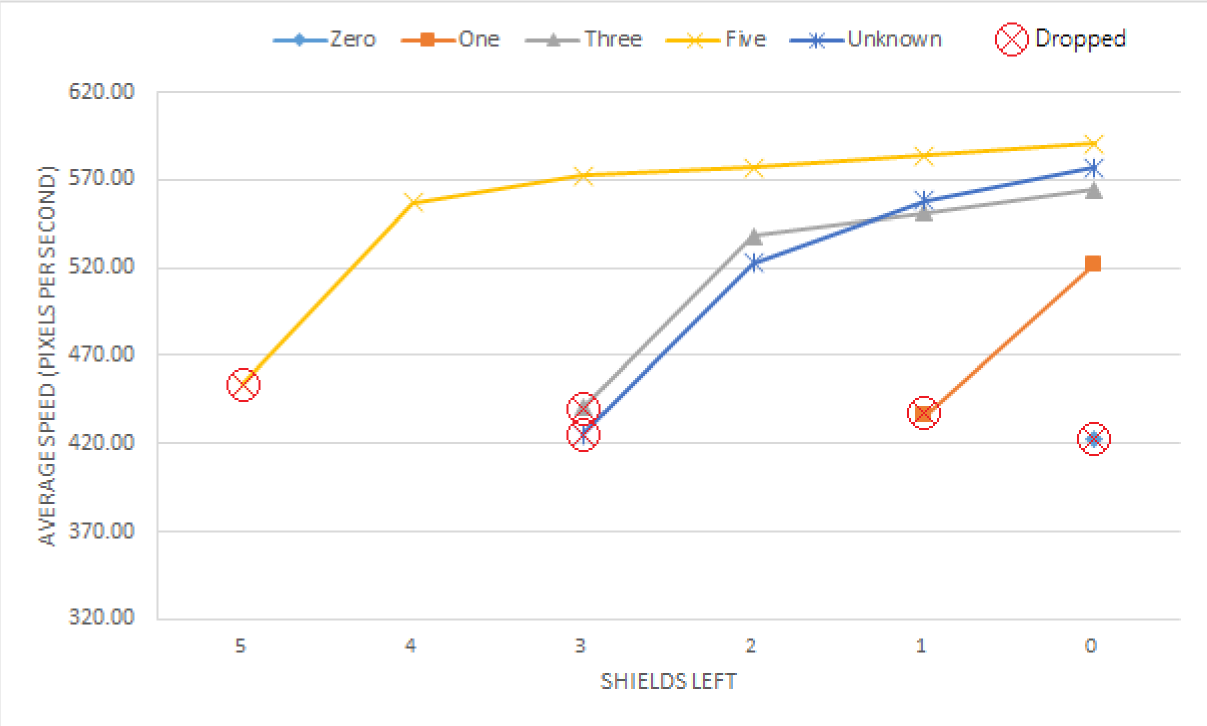

Figure 3. Plot of mean speed per shield situation (line diagram). 320px/s is the minimum speed. Note that one can count the seventeen shield situations in this diagram.

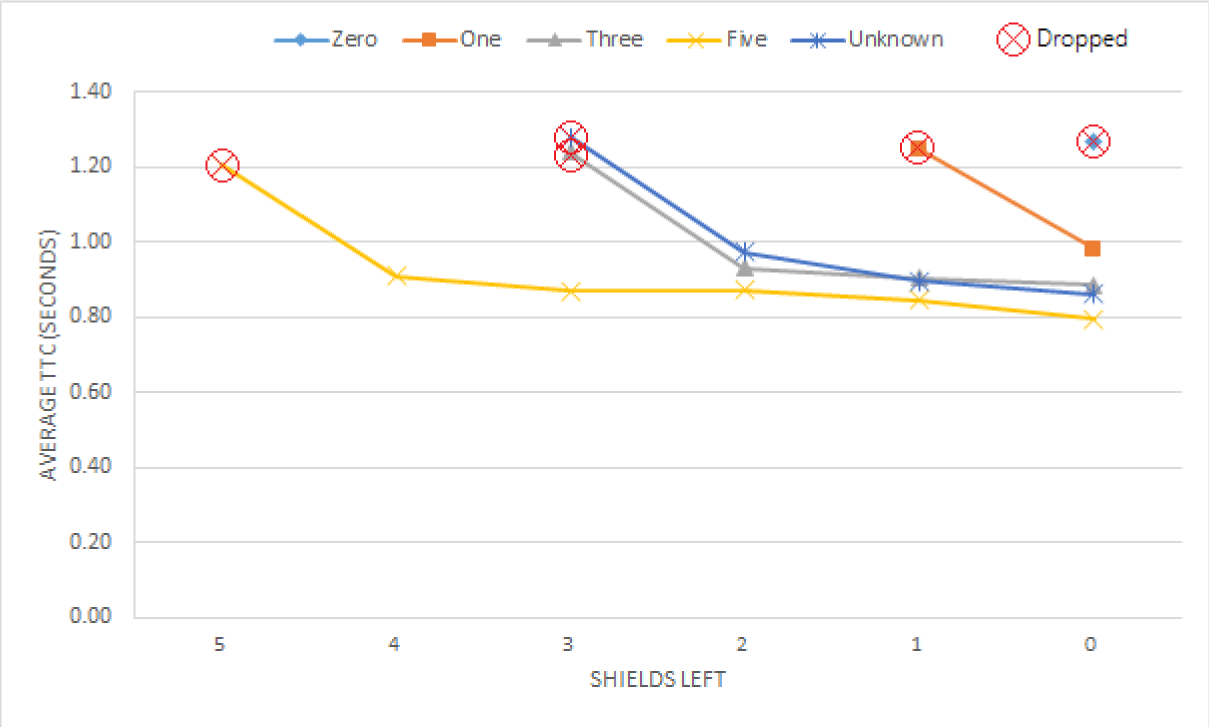

Figure 4. Plot of mean ttc per shield situation (line diagram).

The analysis of the data to pinpoint a homeostatic effect was done on two levels. Firstly, risk-taking behavior averages of shield situations within each condition are compared to each other. Changes in risk-taking behavior are proof of participants adapting to a changing safety situation. Secondly, condition averages are compared to each other.

None of the speed measurements were normally distributed according to the Kolmogorov-Smirnov and Shapiro-Wilk tests (at $p = 0.05$ significance level). However, skewness and kurtosis levels were within acceptable values, with 1.06 being the largest difference from zero. Four out of twelve measurements of ttc were normally distributed according to above tests. Skewness and kurtosis values were well within acceptable values, with 0.73 being the largest difference from zero. Because skewness and kurtosis are not too bad and the F-test is robust against normality with large sample sizes, it is viable to continue further analysis.

The amount of shields that participants had left in a session did not affect their speed nor ttc in the three and five-shields conditions (Table 1). However, when the amount of shields was unknown to the participants, speed and ttc were significantly affected by the amount of shields participants had actually lost. Bonferroni corrected post-hoc pairwise comparisons of the three shield situations in the unknown amount of shields condition were performed. It was found that only the average speed increase from the final shield to having no shields left did not increase significantly (MD = 9.80, p = .18). Ttc only decreased significantly from having two shields and no shields (MD = -0.07, p < .05).

Table 1

Within-subjects effects of shield situation or condition on measurements of risk-taking behavior.

Measurement: average speed

	Sphericity[+]				Within-Subjects Effects				
Condition	Test statistic		df	Epsilon	Test statistic	df1	df2	Sign.	Effect size
	W	X^2			(F)			(p)	(η^2)
3 Shields	.77	37.16*	2	.82[b]	1.27	1.65	240.48	.28	.01
5 Shields	.16	232.77*	9	.50[a]	0.88	1.99	257.27	.42	.01
Unknown	.64	61.45*	2	.74[a]	14.36	1.47	202.94	< .001	.09
Between	.94	10.33	5		3.69	3	474	< .05	.02

Measurement: average ttc

	Sphericity[+]				Within-Subjects Effects				
Condition	Test statistic		df	Epsilon	Test statistic	df1	df2	Sign.	Effect size
	W	X^2			(F)			(p)	(η^2)
3 Shields	.89	17.21*	2	.91[b]	0.33	1.82	265.68	.70	< .01
5 Shields	.64	57.85*	9	.84[b]	1.31	3.35	432.71	.27	.01
Unknown	.99	1.08	2		4.33	2	276	< .05	.03
Between	.87	22.50*	5	.93[b]	3.61	2.79	441.42	< .05	.02

Note. [+] Sphericity is examined for the adjustment of the degrees of freedom. * Significant at the p < .001 level. [a] Greenhouse-Geisser correction. [b] Huynh-Feldt correction.

The kurtosis and skewness values for condition averages of speed and ttc measurements are within acceptable values, respectively differing from zero by 0.76 and 0.69 at most. Only the ttc measurement in the one-shield condition was not significantly different from a normal distribution, according to Kolmogorov-Smirnov and Shapiro-Wilk tests. For the same reasons as stated above, however, it is viable to continue analysis. There was a significant difference between conditions on both risk-taking behavior measurements (Table 1). However, Bonferroni corrected post-hoc pairwise comparisons show that only the one and five-shields conditions differed significantly from each other. In the one-shield condition, participants had lower speed ($MD = 32$, $p < .05$) and higher ttc ($MD = 0.70$, $p < .05$).

Self-efficacy

A scale of self-efficacy in games was developed for this study. The scale measures ones self-efficacy pertaining to how quickly one would be able to understand the mechanics, controls, and objectives and whether they would perform well, be able to complete the game at the highest difficulty level, and perform above average. Reliability analysis showed an excellent scale for self-efficacy with a Cronbach's alpha of 0.93. The squared multiple pairwise correlation between each item was larger than 0.61. The items were summed into a self-efficacy score, which was entered into regression analysis as a predictor of score. There was a significant linear relationship between self-efficacy and score ($R2 = 0.19$, $p < .001$). Skewness (-0.41) and kurtosis (-0.25) values were acceptable, and according to the Kolmogorov-Smirnov ($K\text{-}S = 0.06$, $p > .20$) test the distribution of self-efficacy was normal, albeit not so according to the Shapiro-Wilk ($S\text{-}W = 0.98$, $p < .01$) test. Visual examination of a predicted versus residual scatterplot showed slight heteroscedasticity (Figure 5). The regression model with self-efficacy in games as a predictor of score was found to be significant ($F(1, 165)$ 39.07, $p < .001$). Both the constant ($B = 599.81$, $t(166) = 2.60$, $p < .05$) and the self-efficacy scale ($B = 25.43$, $t(166) = 6.25$, $p < .001$) coefficients were significant predictors of score.

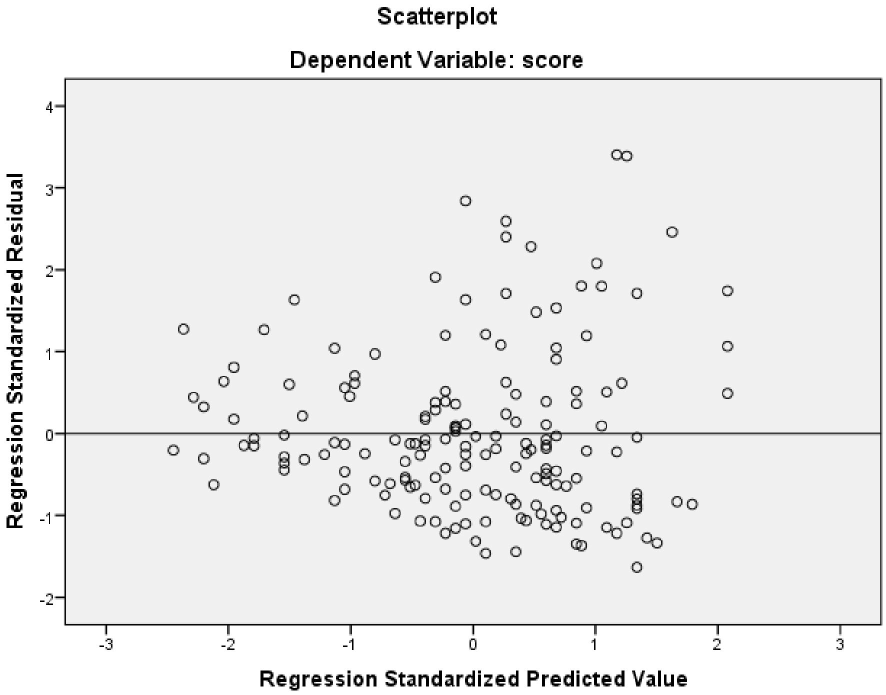

Figure 5. Plot of predicted values versus residual values of score by the regression model with self-efficacy as a predictor (scatterplot).

To find the effect of self-efficacy on risk-taking behavior in the game, the self-efficacy score was added to the above analysis as a covariate. Self-efficacy score had significant positive correlations with speed (.30 < r < .44) and significant negative correlations with ttc (-.38 < r < -.16) measurements. The between-subjects effects of self-efficacy as a covariate of the within-subject effects of condition or shield situation on risk-taking behavior are all significant (Table 2).

Table 2

Between-subjects effects of self-efficacy on measurements of risk-taking behavior in each shield situation or condition.

Measurement: average speed.

Condition	Test statistic (F)	df1	df2	Sig. (p)	Effect size (η^2)
3 Shields	23.58	1	144	< .001	.14
5 Shields	30.21	1	128	< .001	.19
Unknown	29.97	1	136	< .001	.18
Between	31.83	1	156	< .001	.17

Measurement: average ttc.

Condition	Test statistic (F)	df1	df2	Sig. (p)	Effect size (η^2)
3 Shields	24.02	1	144	< .001	.14
5 Shields	23.17	1	128	< .001	.15
Unknown	23.13	1	136	< .001	.15
Between	24.23	1	156	< .001	.13

With self-efficacy in games as a covariate, both speed and ttc changed significantly as shields were lost in the five-shields condition. In the unknown amount of shields condition, the main effect of speed was lost by adding self-efficacy as a covariate, but the effect on ttc was significant (Table 3). There is no interaction effect of self-efficacy and condition on risk-taking behavior.

Table 3

Within-subjects interaction effects of self-efficacy and shield situation or self-efficacy and condition on risk-taking behavior measurements.

Measurement: average speed

Condition	Sphericity[+] Test statistic W	X^2	df	Epsilon	Within-Subjects Effects Test statistic (F)	df1	df2	Sig. (p)	Effect size (η^2)
3 Shields	.78	35.95*	2	.83[b]	1.09	1.66	239.62	.33	.01
5 Shields	.18	216.51*	9	.52[a]	6.28	2.06	263.80	< .01	.05
Unknown	.64	60.67*	2	.73[a]	0.95	1.47	199.71	.37	.01
Between	.94	9.45	5		0.06	3	468	.98	< .01

Measurement: average ttc

Condition	Sphericity[+] Test statistic W	X^2	df	Epsilon	Within-Subjects Effects Test statistic (F)	df1	df2	Sig. (p)	Effect size (η^2)
3 Shields	.89	16.47*	2	.92[b]	1.19	1.84	264.67	.30	.01
5 Shields	.66	53.27*	9	.86[b]	3.85	3.44	440.23	< .01	.03
Unknown	.99	0.98	2		4.20	2	272	< .05	.03
Between	.86	23.22*	5	.94[b]	0.26	2.81	437.72	.84	< .01

Note. [+] Sphericity is examined for the adjustment of the degrees of freedom in the case of absence of sphericity. * Significant at the $p < .001$ level. [a] Greenhouse-Geisser. [b] Huynh-Feldt.

By recoding self-efficacy to make two groups, the effects of self-efficacy as a covariate can be visualized (Figure 6, Figure 7, and Figure 8). Values are recoded so that all self-efficacy scores below 50 are in group 1 and those above 50 in group 2. The two groups are nearly equal, with 72 of 168 participants (43%) in group 1.

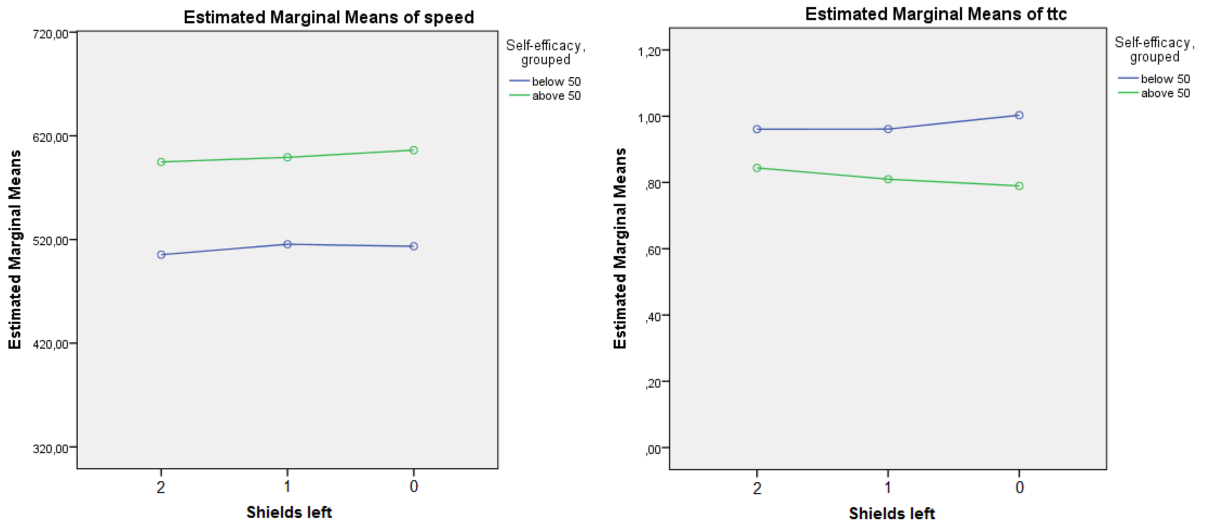

Figure 6. Estimated marginal means of speed (left) and ttc (right) in the three-shields condition, split by self-efficacy group. Values of “shieldspent” are related to shield situation, with 1 being two shields left, 2 being one shield left, and 3 being zero shields left (line diagram). 320px/s is the minimum speed.

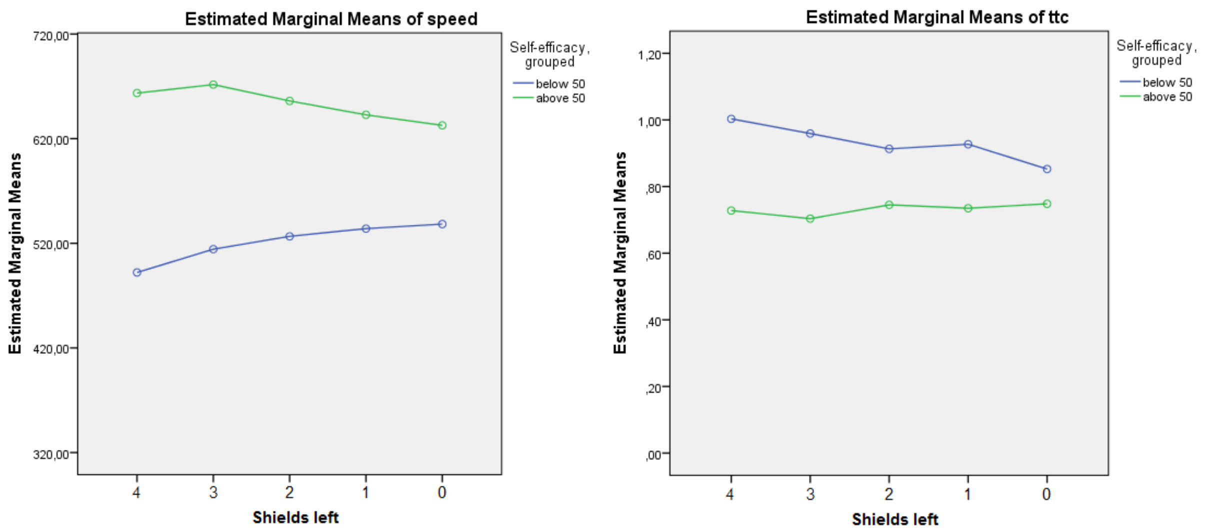

Figure 7. Estimated marginal means of speed (left) and ttc (right) in the five-shields condition, split by self-efficacy group (line diagram). 320px/s is the minimum speed.

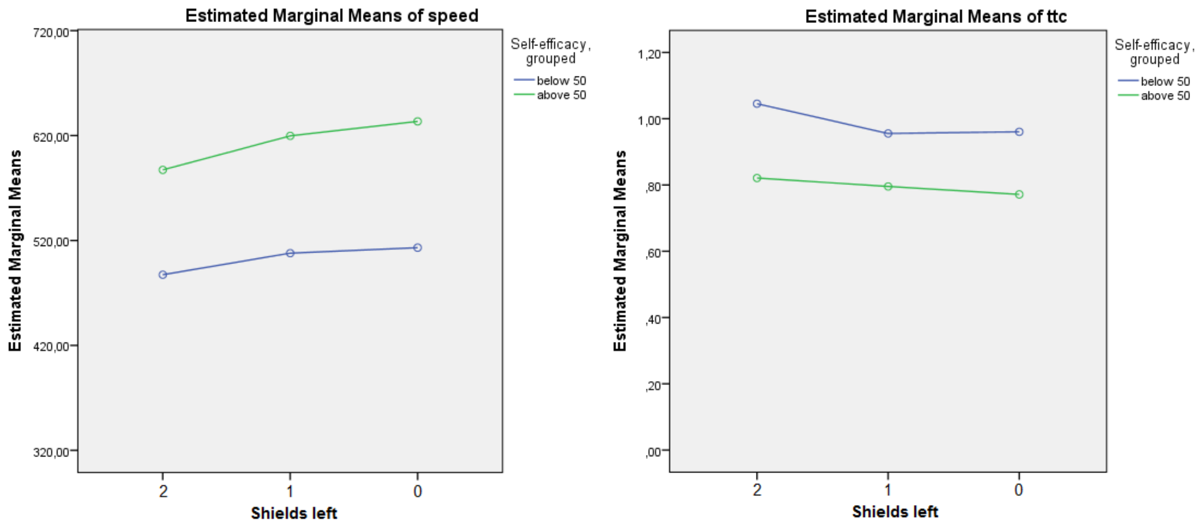

Figure 8. Estimated marginal means of speed (left) and ttc (right) of the unknown-shields condition, split by self-efficacy group (line diagram). 320px/s is the minimum speed.

To visualize the origin of the effects, boxplots of average speed in the five-shields and three-shields condition were created. The boxplots were split by self-efficacy group (Figure 9). In the three-shields condition, the two self-efficacy groups do not have enough differences across shield situations for statistical significance. In the five-shields condition, however, high self-efficacy players fly faster in the beginning and lower their speed to a level that is similar to that of low self-efficacy players.

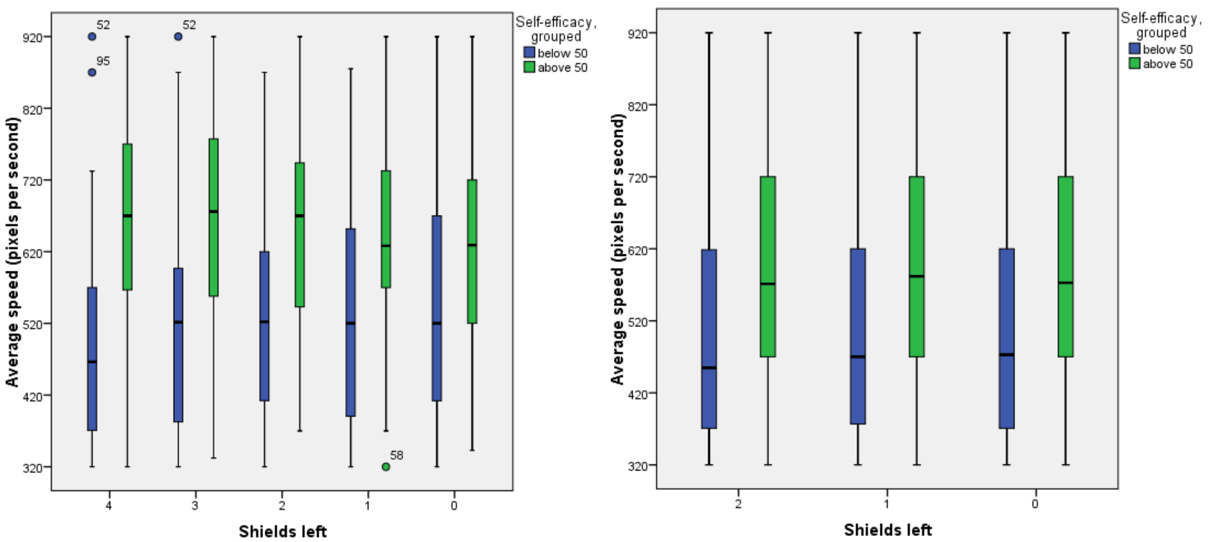

Figure 9. Visualization of the distribution of the data of average speed in every shield situation of the five-shields (left) and three-shields (right) conditions, split by self-efficacy group (boxplot). 320px/s is the minimum speed.

Discussion

There were three analyses in the scope of this study. The effects of shield situation, within each condition, or condition on risk-taking behavior were examined in the first analysis. In the second analysis, the reliability of a new self-efficacy scale and its relationship to performance in the game was examined. Finally, the first analysis was repeated, with self-efficacy added as a covariate. This analysis could find whether there is an interaction of self-efficacy with the effect of the amount of shields left on risk-taking behavior. The results are summarized shortly before they are discussed in further detail.

This study has made several findings. Firstly, there is no direct evidence of risk compensation as players of a game lose their shields and come closer to losing their ship (H1). Secondly, a newly developed self-efficacy scale has good internal reliability, and can significantly predict performance in the game in this sample (H2). Thirdly, self-efficacy had a significant effect on risk-taking behavior (H3). Fourthly, self-efficacy scores were found to interact with the effect of the amount of shields players had left on risk-taking behavior, but only in specific situations (H4). The specificity of these situations suggests that self-efficacy is probably a factor in the perception of risk. In light of these results, the hypotheses that were stated in this experiment cannot be accepted.

Risk-taking behavior

Firstly, there was no proof of participants adapting their risk-taking behavior according to differing safety situations (H1). Only in the unknown-shields condition, the shield situation affected risk-taking behavior. Participants kept increasing, instead of reducing, their risk-taking behavior by increasing speed and reducing their ttc. In the other conditions, however, risk-taking behavior remained constant. While we would have expected the risk taking behavior to go down as players lose shields, some underlying variables in Wilde's (1982a) original model (Figure 2) might be counteracting this, causing the risk-taking behavior to stay at the same level. There is some grounds to discuss that some form of risk compensation is taking place. In its absence, risk taking would have been increasing in every condition. Herein is assumed that risk compensation does not happen or happens to a lesser extent in the unknown shields condition. This is to say that risk taking is some product of underlying variables that increase or decrease risk taking. In the case of this experiment the net result of this product is a constant level of risk taking.

Another explanation for the absence of reduced speed is that instead of gaining points by staying alive longer participants may have wanted to keep an increased speed to maximize the amount of points they could get in their last life. Otherwise, the perceptual difference between having two shields or zero shields left may be too small for any participant to judge the situation as being more dangerous. In addition, participants might judge that flying faster is not more dangerous as they were getting used to the speed. In conclusion, the risk-taking behavior is matched to the target level of risk relatively quickly and the perceived risk somehow remains constant even though participants lose shields.

Secondly, condition had a significant effect on risk-taking behavior. In pairwise comparison, however, only the one-shield condition differed from the five-shields condition. In the one-shield condition participants played less risky by flying their ship slower and dodging incoming meteors sooner. The three-shields and unknown-shields conditions were somewhere in between the one-shield and five-shields conditions, albeit not significantly different from either. This finding does suggest that participants actually played more carefully in a risky situation than in a safe situation, and thus at least compensate for added or reduced risk.

Self-efficacy and risk-taking behavior

The self-efficacy scale had excellent internal reliability as a scale in this sample and was a significant predictor of score in the game (H2). Validation by comparing this scale to other scales that measure the same or similar constructs is warranted for future use. Further analysis should consider whether six items are appropriate and which aspects of self-efficacy in games are not covered.

There was a between-subjects effect of self-efficacy on speed and ttc in all shield situations and conditions (H3). Risk-taking behavior was higher for participants that had higher self-efficacy. Additionally, self-efficacy did interact with shield situation on risk-taking behavior between shield situations within the five-shields condition, and ttc only in the unknown-shields condition. Higher self-efficacy scores of participants led to more risk-taking behavior per shield lost than lower self-efficacy scores. The absence of this finding in the three-shields condition is puzzling at least. Confident players perhaps judged that having three shields is not as safe as having five shields. These players will only increase their risk-taking behavior compared to unconfident players if the situation is judged as safe. The effect of self-efficacy on risk-taking behavior seems to diminish as the shield situation becomes more dangerous. This notion is confirmatory of the hypothesis that, in fact, self-efficacy does play a role in the perception of risk in the process of risk compensation.

Between conditions, however, there was no interaction with self-efficacy. Players with higher self-efficacy scores took more risk than players with lower self-efficacy scores in every condition. However, this increased risk-taking was equal across all conditions, hence the absence of an interaction effect. While this finding speaks against the hypothesis (H4), it does not refute that self-efficacy plays a role in risk perception. Averaging risk-taking behavior within conditions may be a bad way to examine risk compensation because it summarizes much meaningful data.

Limitations

Proving risk homeostasis theory is a bridge too far for this research design. The game might not give enough feedback on the risk-taking behavior of the participant. In the real world, many social elements give feedback on behavior. Getting a dent in your car is not only reducing the aesthetics of the car, but you will have to explain to friends and family what behavior had led to it. If players could compare their score to other players there would be more and better feedback. Then, players might more thoroughly judge their expected benefits and losses from risky or safe behavior, which is an important part of risk homeostasis theory as discussed in the introduction. Additionally, there might not be enough time for players to learn the costs and benefits of their behavior in the game. The effects of more in-depth training sessions should be examined. Improving feedback might also help to increase the effect sizes. The resulting effect sizes from the analyses in the current study are quite small. Another way that might help to increase effect sizes is dramatizing the effect of hitting a meteor and putting more emphasis on the amount of shields that players had.

The exclusion of the first shield that participants lost in every condition is unfortunate, but no deal-breaker. One could argue that players will always treat their first shield with extra care, thusly explaining the lower speed. This would have been an interesting finding, albeit very unlikely. To avoid this from happening in future research there are two solutions. Either the first few seconds of data collection should be excluded from analysis or the speed should be set randomly at the start. While the first offers a challenge pertaining to choosing a point in time to cut off data, the second may introduce unnecessary noise in the data. A third option is to set the speed of the ship to the overall average that has been found in this and previous studies, allowing participants to slow down or speed up as they please.

This study contained some other source of bias in measurement. Participants that played more risky actually lost all their shields, while participants that played less risky made it to the end without losing their ship. Participants that made it to the end by flying for four minutes have missing data for the shields they did not lose. Therefore, the average risk-taking behavior is lower when there are more shields and higher when there are less, relative to what would have been found if there had been no missing data. Players that made it to the end and thus had missing data for some of the shields they did not lose should be excluded or compared to players that lost their ship. This way we can know whether there was actually a bias in the measurement. However, removing these cases from analysis actually means removing all cases with a safe playstyle, introducing a range of other limitations. Alternatively, the game could be gradually increased in difficulty over time until impossible to complete, or there should be certain points where players definitely lose a shield, so that even safe players have no missing data. Fair warning should be issued for frustrated participants due to making them play an unbeatable game.

As shortly discussed above, participants might be taking increasing amounts of risk because they are starting to feel familiar with the game, counteracting the effects of finding themselves in a more dangerous situation. The training round might have been too short to familiarize participants with the game. An argument that alleviates the impact of this flaw is that the novelty of a situation is actually inherent to risk perception. Nonetheless, it may prove interesting to further study these effects in detail to increase effect-size. In reality, nearly all car drivers have had sufficient experience and training. The unfamiliarity that players had with the game could be controlled for by comparing risk-taking behavior at the first few minutes to a later point in time. The effect of differing lengths of practice sessions or the introduction of in-depth training sessions should be examined.

Another source of bias in measurement concerns the calculation of ttc. The data did not exclude ttc data where the meteor was hit, causing noise in the data so that averages are lower than they should be. The ttc should have only been logged for meteors that were actually dodged.

Finally, self-efficacy was measured after participants played the game. Therefore, a judgement of their own performance in the game could be a large latent factor in the explained variance of the self-efficacy score. However, there are several arguments for choosing to present the questionnaire after the game. Firstly, the questions were carefully chosen so they have very general meaning, using wording such as “in a game”. Secondly, there was no way for participants to have an objective view of their final score, nor compare their performance to others. Lastly, the priming effects of asking people about their skills before they play the game could have worse effects than. Optimally, the scale is tested by presenting it before or after a game and examining the differences in the relationships between performance and self-efficacy.

Future research

The unknown shield condition is possibly the best reflection of reality. There is hardly a factual and quantifiable representation of the risk of a situation in reality. Hitting a meteor in the unknown-shields condition will instill the feeling that the situation has probably become more dangerous. The comparison of behavior in this condition to that in which the amount of shields is known led to some interesting results, as discussed above.

The scale for self-efficacy in games was reliable in this sample and had a significant moderate predictive value on performance in the game. While further testing of the scale in other samples and validation through correlation with related constructs is warranted, initial results are promising.

A complicated theory with so many factors warrants more research. Some suggestions for future research were described in the discussion of the results above. However, most of these are related to technical limitations or constraints of this specific game and research design. Other, more general, elements that warrant further research are for example the effects of experience. Perhaps, longer training sessions leads to different risk-taking behavior. Then, the difficulty of the task may play a large role. It is not inconceivable that there is an interaction effect between skill and task difficulty on risk-taking behavior. Another major factor that has not been investigated thus far is motivation and effort.

Naturally, many other factors are related to risk-taking behavior, which could be studied in detail. However, the main goal of such research is to prove risk homeostasis theory in all humans. The main goal is not trying to find its occurrence in very specific situations. Sadly, the results to prove the existence of a risk homeostasis process is sparse in this study. Future research should foremost focus on increasing the power of the tests that have been used, as opposed to finding more specific circumstances in which homeostatic processes may emerge.

To conclude, this study has not been able to show a risk homeostatic process. Only a strong decline in risk-taking behavior as players lost shields would justify the conclusion that risk homeostatic processes were in effect. Some findings suggest risk compensation in certain situations. An interaction effect with self-efficacy was found. The self-efficacy scale developed for this study was shown to have good internal reliability and predictive quality of score and risk-taking behavior.

References

Adams, J. G. (1988). Risk homeostasis and the purpose of safety regulation. *Ergonomics, 31*(4), 407-428.

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Evans, L. (1986). Risk homeostasis theory and traffic accident data. *Risk Analysis*, *6*(1), 81-94.

Glendon, A. I., Hoyes, T. W., Haigney, D. E., & Taylor, R. G. (1996). A review of risk homeostasis theory in simulated environments. *Safety science, 22*(1), 15-25.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Llewellyn, David J., & Sanchez, Xavier. (2008). Individual differences and risk taking in rock climbing. *Psychology of Sport & Exercise, 9*(4), 413-426.

Sorrentino, R. M., Hewitt, E. C., & Raso-Knott, P. A. (1992). Risk-taking in games of chance and skill: Informational and affective influences on choice behavior. *Journal of Personality and Social Psychology, 62*(3), 522.

Trepte, S., & Reinecke, L. (2011). The pleasures of success: Game-related efficacy experiences as a mediator between player performance and game enjoyment. *Cyberpsychology, Behavior, and Social Networking, 14*(9), 555-557.

Wilde, G. J. (1982a). The theory of risk homeostasis: implications for safety and health. *Risk analysis*, *2*(4), 209-225.

Wilde, G. J. (1982b). Critical issues in risk homeostasis theory. *Risk Analysis*, *2*(4), 249-258.

Wilde, G. J. S. (1988). Risk taking in psychomotor and cognitive tasks as a function of probability of loss, skill and other person-related variables. In *Road user behaviour: Theory and research* (pp. 120-126). Van Gorcum, Assen The Netherlands.

Wilde, G. J. (1998). Risk homeostasis theory: an overview. *Injury Prevention, 4*(2), 89-91.

Appendix A

Self-efficacy in videogames (SE videogames).

I can perform well in a game

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)

I can understand and use controls effectively for a game quickly

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)

I can understand the mechanics of a game quickly

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)

I can complete the hardest difficulty setting of a game

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)

I can perform above average when playing against other players

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)

I can quickly understand the objective in a game

0 (cannot do at all) – 50 (moderately can do) – 100 (highly certain can do)