Bayesian Highest Density Intervals of Take-Over Times for Highly Automated Driving in Different Traffic Densities

Moritz Körber, Jonas Radlmayr, Klaus Bengler
Technical University of Munich – Chair of Ergonomics

Current human factors research on automated driving aims to ensure its safe introduction into road traffic. Although informative results are crucial for this purpose, most studies rely on point estimates and dichotomous reject-nonreject decisions that have been declared obsolete by more recent statistical approaches like *new statistics* (Cumming, 2014) or Bayesian parameter estimation (Kruschke, 2015). In this work, we show the objective advantages of Bayesian parameter estimation and demonstrate its abundance of information on parameters. In Study 1, we estimate take-over times with a relatively uninformed prior distribution. In Study 2, the resulting posterior distributions of Study 1 were then used as informed prior distributions for interval estimations of mean, standard deviation and distribution shape of take-over time in different traffic densities. We obtained 95 % credible interval widths between 490 ms and 600 ms for mean take-over times, depending on the condition. These intervals include the 95 % most probable values of the mean take-over time and represent a quantification of uncertainty in the estimation. Given the data and the experimental conditions, a take-over requires most likely 2.51 seconds [2.27, 2.76] when there is no traffic, 3.40 seconds [3.11, 3.71] in medium traffic and 3.50 seconds [3.21, 3.78] in high traffic. Bayesian model comparison with Bayes Factor is discussed as an alternative approach in conclusion.

CURRENT STATISTICAL METHODS IN RESEARCH ON AUTOMATED DRIVING

The introduction of automated vehicles to road traffic is accompanied by discussion and research on technical, legal, and human factors aspects. The aim of human factors research activity in this domain is to predict the interaction between human and vehicle automation in order to detect any kind of limitation of human performance, error proneness, potential for misuse or similar potential threats for a safe use. A frequent research topic is the estimation of the take-over time – a certain type of response time which is defined as the time between the automated vehicle's request from the driver to regain vehicle control and the first conscious manual driver input (Gold & Bengler, 2014). These estimates have to be as precise as possible in order to arrive at valid conclusions about the safety of automated driving. Take-over times have already been estimated in past publications, however, these estimates are unfortunately mostly reported and interpreted as point estimates (Gold, Damböck, Lorenz, & Bengler, 2013; Radlmayr, Gold, Lorenz, Farid, & Bengler, 2014; Zeeb, Buchner, & Schrauf, 2015). Although the standard deviation gives information about the dispersion, point estimates give no information about the precision of an estimation. Both point estimates and blind reliance on a dichotomous reject-nonreject decision strategy based on null hypothesis significance testing (NHST) have been deemed as uninformative, flawed, ineffective for scientific progress and thus obsolete by more recent approaches based on interval estimation combined with effect sizes (referred to as *new statistics*; Cumming, 2014) and Bayesian parameter estimation (Kruschke, 2015). Although other research domains are already transitioning to these approaches (Kruschke, Aguinis, & Joo, 2012), human factors research on automated driving has yet not implemented them. This work, therefore, discusses two different approaches to interval estimation and applies Bayesian parameter estimation to take-over times.

FREQUENTIST AND BAYESIAN PARAMETER ESTIMATION

In frequentist inference, the precision of an estimation is expressed in form of a confidence interval (CI). A confidence interval is defined as "an interval or range of plausible values for some quantity or population parameter of interest" (Cumming & Finch, 2001, p. 533). This interval is centered on the best point estimate while its width represents the imprecision of the estimate, which is caused by sampling error and measurement error. Its width is also determined by a chosen *confidence level* c, with higher desired confidence resulting in wider intervals (other parameters being equal). Since experimental results with samples are merely estimations of population parameters, CIs reflect the view of results as estimates better than point estimates, because CIs report not only effect size but also precision (Cumming, 2014). The interval represents a set of plausible values for a parameter (e.g. population mean) and values outside of the interval are deemed implausible. If assumptions are met, an individual estimation error – the difference of the sample mean from the true mean of a population – is not greater than the *margin of error* (MOE) of a CI at a chosen confidence level c, e.g. 95 % (Cumming, Fidler, Kalinowski, & Lai, 2012). Because of these advantages over NHST, Cumming (2014) promotes the use of CIs in his data analysis approach called the *new statistics* and encourages a paradigm shift in data analysis towards it. However, probabilities that are associated with CIs are subject to the same assumption as p-values, namely that the only source of imprecision is sampling error, which is doubtful to be fulfilled in empirical studies (Kline, 2013). CIs represent a frequentist method of inference and therefore have to be interpreted in frequentist terms: A $(1 - \alpha)$ % confidence interval is an interval from a notional infinite sequence of random samples that includes the unknown, fixed parameter θ $(1 - \alpha)$ % of the times (Steiger & Fouladi, 1997). This follows from the frequentist interpretation of probability as the asymptotic relative frequency of an event in infinitely repeated trials under identical conditions and error only by chance (Brandstätter & Kepler, 1999; Reichardt & Gollob, 1997). Because the population parameter is fixed, the conclusion that a 95 % CI has a 95 % probability to include it is not correct. Frequentism allows no such statement for any single certain CI, thus, the parameter is either in the interval or not. Because of this, confidence intervals yield no distributional information and parameter values in the center of CIs are not more probable than values at its limits (Kruschke & Liddell, 2015). Distributions that are sometimes superimposed on CIs (Cumming & Fidler, 2009) are hypothesized sampling distributions of the parameter's sample estimator. Because of this, they suffer from the same flaw as *p*-values since both are determined by the researcher's stopping and testing intentions. Depending on the choice of them, one can create many different CIs out of a single data set (Kruschke & Liddell, 2015). In fact, Morey et al. (2015) show that CIs do not necessarily have any of the attributed properties, and may lead to irrational inferences. Bayesian statistics circumvent this limitation, firstly, because probability is interpreted not as an asymptotic relative frequency but as belief, plausibility or state of knowledge. Secondly, a population parameter is not seen as a fixed, true value, but as a random variable with its own distribution based on plausibility and the state of knowledge. Bayesian parameter estimation represents the process of updating a prior belief or state of knowledge on a parameter value by experimental data. This results in a posterior probability distribution (referred to as *posterior*) that depicts the likelihood of a parameter value given the experimental data. This distribution provides the expected value of the distribution as a point estimate and, at the same time, reflects the uncertainty of this estimate (Kline, 2013). Intervals of this posterior distribution are called *credible intervals* or *highest density intervals* (HDI). Therefore, a Bayesian 95 % credible interval implies what is often incorrectly attributed to CIs: it includes the parameter with a probability of 95 % (Castro Sotos, Vanhoof, van den Noortgate, & Onghena, 2007). That is because the posterior represents an actual probability distribution of the parameter value, given the data, and the HDI is the constructed so that it contains the 95 % of this distribution. Parameter values in the center tend to have higher credibility than parameter values at the limits. By the implementation of prior knowledge into the estimation, the Bayesian view on probability thus takes into account that not all parameter values or hypotheses are equally plausible or likely. Prior knowledge is used by the researcher to state his prior hypothesis on the parameter value before data are collected or analyzed. The experimental data then update this prior belief. The relationship between the prior hypothesis and the experimental results is reasonable: exotic or implausible hypotheses require strong evidence to be supported and the prior becomes less and less impactful the more empirical evidence is collected. Imprecise data have less impact on the posterior in relation to the prior than precise data. The same is true for the prior: the more informed the prior is formulated, the more weight it has on the posterior. Confidence intervals and Bayesian intervals only yield the same information if the chosen prior distribution is a flat uninformed prior, i.e. no prior knowledge or belief is available and the posterior distribution is a result solely from sample data (Reichardt & Gollob, 1997). It is rarely the case that a researcher has no prior belief about his hypotheses and thus a CI is not optimal for parameter estimation if prior knowledge exists. Usually, an informed prior can be formulated based on prior beliefs, knowledge or previously collected empirical data. For the purpose of parameter estimation, the prior is then updated by the obtained experimental data and this process results in extended, updated knowledge, the posterior, which is the probability distribution of the parameter value given the obtained data. The estimates of the posterior distribution are a weighted combination of prior distribution and the empirical evidence (data) while the weight is determined by each source's precision (Kline, 2013). The aim of this article is to perform such a Bayesian interval estimate on take-over times in different traffic conditions by creation of a 95 % HDI based on prior experimental data. We use a previously conducted study (Study 1) as prior knowledge on the central tendency, dispersion and shape of the take-over time that is estimated by the data of Study 2.

STUDY I

This study was used to formulate a prior distribution for study 2 that is informed and generated by empirical data. In order to generate a plausible informed prior distribution, we chose an experiment that had similar conditions. A more detailed description of the experiment can be found in Radlmayr et al. (2014); therefore, we only give a brief description of the experimental setup here.

Sample and experimental setup

A total of n = 16 participants, 3 (19 %) female and 13 (81 %) male, were analyzed in this study. Mean age of the sample was M = 34.88 years (SD = 11.15) with a range from 19 to 58 years. Participants possessed their driver's license for a mean duration of M = 20.60 years (SD = 9.12) and the majority of them (10; 62.50 %) indicated that their annual mileage was between 5,000 km and 10,000 km per year. 12 (75 %) participants have already taken part in a driving simulator experiment before. No participant had experience with highly automated driving (HAD) before.

Participants drove highly automated (Gasser, 2012; Level 3 in NHTSA, 2013) on a six-lane highway for the whole experimental track. HAD allows the driver to engage in a non-driving-related task and to simulate this activity, the participants had to solve a cognitive task (2-back task; Reimer, Mehler, Wang, & Coughlin, 2010) and the Detection Response Task (DRT; Conti, Dlugosch, Vilimek, Keinatz, & Bengler, 2012) during the drive prior to the take-over process. The take-over request (TOR) was given due to an obstacle in form of a car crash with two stationary vehicles on participant's lane. A seven seconds (= 233 m at 120 km/h) take-over time budget was provided. We analyzed two conditions: with (*traffic*) and without (*no traffic*) other surrounding road users.

Generation of a posterior distribution of take-over times

We conducted the parameter estimation based on scripts by Kruschke (2015) implemented in the statistical computer software R. JAGS (Plummer, 2003) was used as Markov Chain

Monte Carlo (MCMC) method to approximate the posterior distribution. With this method, the integral for the cumulated evidence, p(Data), does not have to be explicitly calculated. A Markov chain represents a stochastic process which stepwise samples a large number of combinations of credible parameter values from the posterior distribution that each accommodates both data and prior distribution. For every discrete step, the likelihood of this parameter combination is calculated. The sequence of steps is probabilistic such that the likelihood of a combination scales with its frequency of appearance in the chain. Repeating this step a large number of times, the evaluation of the chain results in a high resolution approximation of the continuous posterior distribution (see Figure 1 for an example). For the estimation of the take-over time y, we chose a generic, relatively uninformed and robust prior distribution that was only minimally informed by the scale of the data. The mean μ was estimated with a normal prior distribution that was adjusted to the descriptive values of the sample data. σ was estimated with a flat uninformed prior with boundaries adjusted to the sample standard deviation. To take potential outliers into account, the take-over time y was assumed to follow a heavy-tailed t-distribution instead of a normal distribution. Therefore, the normality parameter ν, which determines its tails and thereby its shape, was estimated as well.

$\mu \sim \text{normal}(\text{mean}(y), 1/(\text{SD}(y)\ 100)^2)$
$\sigma \sim \text{uniform}(\text{SD}(y)/1000, \text{SD}(y)\cdot 1000)$
$\nu \sim \text{exponential}(1/30.0)$
$y \sim t(\mu, 1/\sigma^2, \nu)$

The distribution of the take-over times was tested for normality graphically (Q-Q-Plot, histogram) and with Kolmogorov-Smirnov-tests with Lilliefors correction ($p_{\text{traffic}} = .200$, $p_{\text{no traffic}} = .012$). Besides one outlier in *no traffic*, normal distribution was given. The obtained MCMC chains have been successfully checked if they meet the quality criteria representativeness, accuracy, and efficiency (Kruschke, 2015). The resulting posterior distributions were then used as an informed prior for the parameter values of Study 2 in order to generate a more precise posterior distribution of the take-over times.

STUDY II

A more detailed description of the experiment can be found in (Körber, Gold, Lechner, & Bengler, 2016), therefore, we only give a brief description of the experimental setup here.

Sample and experimental setup

A total of 72 participants were analyzed, 14 (19.4 %) female and 58 (80.6 %) male. Mean age was $M = 44.97$ years ($SD = 22.16$) with a range from 19 to 79 years. Participants held their license for a mean of $M = 27.48$ years ($SD = 22.37$). Annual mileage per year was between 5,000 km and 20,000 km for the majority of participants (41; 56.94 %). The high standard deviation of age is caused by the initial sampling of participants into a young and older age group. However, since confidence intervals for take-over times overlapped in the analysis between 31 % and 71 % with a mean difference of 254 ms between the groups, both groups were merged into a single sample. While 23 (31.90 %) participants already were experienced with a driving simulator, only 6 participants had taken part in a study on HAD yet.
In this experiment, three traffic density (TD) conditions were simulated in a within-subject design, i.e. each participant experienced a take-over situation in each TD condition. In *Zero TD* (TD 0), no other vehicles were on the road at the moment of the TOR. For the other two conditions, *Medium TD* (TD 10) and *High TD* (TD 20), traffic density was manipulated by the number of other vehicles and their constant and equal distance to the ego vehicle The prior for each traffic condition was informed by the posterior of the corresponding traffic condition in Study 1 (see Table 1).

Table 1: The posterior distributions of Study 1 allocated as prior distributions for Study 2

Posterior		Prior
Study 1 no traffic	⇒	Study 2 no traffic
Study 1 traffic	⇒	Study 2 medium traffic
	⇒	Study 2 high traffic

A hands-free cell phone conversation was simulated by means of the 20 questions task (TQT; Merat, Jamson, Lai, & Carsten, 2012) for half of the participants since this will be a common use case for HAD (Kyriakidis, Happee, & Winter, 2015). The confidence intervals of take-over times of each group overlapped by 75.34 % and 90.03 %, therefore we did not distinguish between both groups in the further analysis. Participants drove highly automated (Level 3 NHTSA, 2013) on a six-lane highway (three lanes in each direction) at a speed of 120 km/h for the whole experimental track. The system limits were each represented by a broken down vehicle on the participant's current lane. It suddenly appeared 233 meters ahead on a straight stretch and an auditory TOR was given seven seconds before the ego vehicle would have collided with the obstacle.

Results: Estimation of a 95 % HDI of take-over time

In condition TD 0, one participant did not notice the TOR and therefore had to be removed from further analysis. The posterior distribution of the parameter values of Study 1 have been used to build an informed prior distribution for the parameter values in Study 2. The same set of parameters was estimated for the three TD conditions. Each prior for μ is a normal distribution with a mean and standard deviation taken from the posterior of Study 1. The prior of σ as well as ν are each given a gamma distribution with each shape and rate (derived from the desired mode and standard deviation of the gamma distribution) based on the posterior of Study 1. A gamma distribution was chosen because it only contains positive values and can implement long tails. The prior for the take-over times y is again a t-distribution.

$\mu \sim \text{normal}(\text{mean}(\mu_{\text{Study1}}), \text{SD}(\mu_{\text{Study1}}))$
$\sigma \sim \text{gamma}(\text{shape}(\sigma_{\text{Study1}}), \text{rate}(\sigma_{\text{Study1}}))$
$\nu \sim \text{gamma}(\text{shape}(\nu_{\text{Study1}}), \text{rate}(\nu_{\text{Study1}}))$
$y \sim t(\mu, 1/\sigma^2, \nu)$

The distribution of the take-over times was again tested for normality graphically (Q-Q-Plot, histogram) and with Kolmogorov-Smirnov-tests with Lilliefors correction ($p_{TD\ 0}$ = .011, $p_{TD\ 10}$ = .200, $p_{TD\ 20}$ = .200). The deviation of TD 0 from normality was caused by one outlier and is seen as non-critical at a sample size of n = 71. The resulting posterior distributions central tendencies and HDIs are shown in Table 2–4.

Table 2: Results of the parameter estimation for TD 0.

	Mean	Median	Mode	HDI Lower Limit	HDI Upper Limit
μ	2.51	2.51	2.53	2.27	2.76
σ	0.92	0.92	0.93	0.75	1.11
ν	27.73	18.69	7.99	2.11	83.03

Table 3: Results of the parameter estimation for TD 10.

	Mean	Median	Mode	HDI Lower Limit	HDI Upper Limit
μ	3.40	3.40	3.42	3.11	3.71
σ	1.25	1.25	1.24	1.05	1.45
ν	46.58	38.50	22.41	4.25	108.64

Table 4: Results of the parameter estimation for TD 20.

	Mean	Median	Mode	HDI Lower Limit	HDI Upper Limit
μ	3.50	3.50	3.52	3.22	3.79
σ	1.18	1.18	1.17	1.00	1.38
ν	45.71	37.30	21.66	4.11	109.06

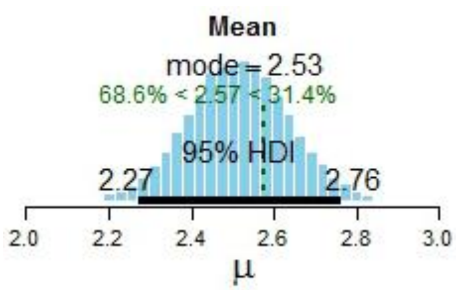

Figure 1: Exemplary posterior distribution for μ in TD 0

Table 5: Comparison between confidence intervals (CI) and highest density intervals (HDI).

	CI Lower Limit	CI Upper Limit	HDI Lower Limit	HDI Upper Limit
TD 0	2.35	2.83	2.27	2.76
TD 10	3.12	3.73	3.11	3.71
TD 20	3.25	3.82	3.22	3.79

Table 6: Comparison between confidence intervals (CI) and highest density intervals (HDI) with n = 15.

	CI Lower Limit	CI Upper Limit	HDI Lower Limit	HDI Upper Limit
TD 0	2.42	3.52	2.53	3.46
TD 10	2.69	3.84	2.66	3.81
TD 20	2.57	3.94	2.59	3.84

DISCUSSION

The purpose of this work was to create a 95 % highest density interval (HDI) for take-over times based on prior experimental data. The results of the parameter estimation offer a detailed interval estimation of take-over times and of the effect of different traffic densities. In addition to a point estimate for the mean take-over time, the uncertainty of the estimation is quantified and a range of the most probable values including their probability distribution is given. The mean take-over time lies within the HDI with a probability of 95 % and the probability given the data of a certain take-over time can be determined. The resulting intervals, shown in Table 2–4, are only marginally smaller than the corresponding CIs (see Table 5), but possess the aforementioned advantages over them in interpretability and abundance of information. A CI relies on a hypothetical sampling distribution which is not a probability distribution. In contrast, the HDI contains the 95 % most probable values of the mean take-over time and offers the probability distribution of interest, i.e. the probability of a take-over time value given the experimental data (= p(take-over time|Data)); whereas the CI only provides the probability p(Data|take-over time). Also, the HDI contains more information at the same size because values in the center have higher credibility than values at the limits, whereas each value is equally likely in a CI. An informed prior makes stronger predictions and therefore leads to more precise estimations if the obtained data match the prior. We reduced the sample size to n = 15 by randomly sampling from the whole sample and re-calculated CIs as well as HDIs (see Table 6). In this case, the HDI is up to 15 %/170 ms smaller than the CI, which is a considerable period in time-critical situations like a take-over. Furthermore, the established posterior distribution provides a prior for future analyses of other researchers. This work also provides practitioners an interval of the most probable mean take-over time values, including their probability distribution and quantified uncertainty. The design and functionality of the vehicle automation can be refined based on this data.

It should be taken into account in the interpretation of the data that Study 1 and Study 2 are very similar experiments, yet the conditions differ slightly. Therefore, the prior distributions for Study 2 did not match the experimental setting perfectly. However, with a sample size like that in Study 2, the prior distribution has less influence on the posterior distribution than the empirical evidence. Future research should not only analyze estimates for means but should also consider the boundaries (e.g. 5 % and 95 % quantile) of the posterior distribution to ensure a safe use of automated driving for the whole population (Körber & Bengler, 2014).

This is the first paper to the authors' knowledge that uses Bayesian parameter estimation in a take-over study, but more and more research domains transition to Bayesian data analysis (Kruschke et al., 2012). We encourage this step and evaluate this method as an addition of another tool for data analysis. The scope of this work was an interval estimation of the parameter take-over time. If a research question requires a hypothesis test, HDIs can further be used for this purpose in the same way as CIs (Kruschke, 2015). However, it is also possible to test hypotheses by comparison of a null model (H_0) with an alternative model (H_1). The comparison is expressed in a likelihood ratio called *Bayes Factor* (BF), which is the ratio between the probability of the data given the null model and the probability of the data given the alternative model. In other words, it expresses if the obtained data are more compatible with a null hypothesis or an alternative hypothesis (Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2015). Usually, the null model states that only a null value (e.g. 0) is possible and uses a prior that allocates all credibility on it. Complementary, the alternative model states that it's possible to attain a broad range of other values as well and uses a prior that allocates credibility over many parameter values (Rouder, Speckman, Sun, Morey, & Iverson, 2009). When compared to NHST, this approach allows sequential testing, a default BF always approaches the correct boundary with increasing sample size, it quantifies evidence both for H_0 and H_1 and allows gradual interpretation of evidence instead of a forced all-or-none decision (Schönbrodt et al., 2015). However, the danger of dichotomous decision making is also prevalent in this approach. The choice for an approach depends on the research question: If it is of interest whether a null model is more or less credible than a specific alternative model, a model comparison is the best choice. If one strives for an estimation to be as exact and informative as possible or the relative credibility of all candidate parameter values is of interest, parameter estimation should be used (Kruschke, 2011).

REFERENCES

Brandstätter, E., & Kepler, J. (1999). Confidence intervals as an alternative to significance testing. *Methods of Psychological Research Online*, *4*(2), 33–46.

Castro Sotos, A. E., Vanhoof, S., van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98–113.

Conti, A., Dlugosch, C., Vilimek, R., Keinatz, A., & Bengler, K. (2012). An Assessment of Cognitive Workload Using Detection Response Tasks. In N. A. Stanton (Ed.), *Advances in human factors and ergonomics series. Advances in human aspects of road and rail transportation* (pp. 735–743). Boca Raton, FL: CRC Press.

Cumming, G. (2014). The new statistics: why and how. *Psychological Science*, *25*(1), 7–29.

Cumming, G., & Fidler, F. (2009). Confidence Intervals: Better Answers to Better Questions. *Zeitschrift für Psychologie / Journal of Psychology*, *217*(1), 15–26.

Cumming, G., Fidler, F., Kalinowski, P., & Lai, J. (2012). The statistical recommendations of the American Psychological Association Publication Manual: Effect sizes, confidence intervals, and meta-analysis. *Australian Journal of Psychology*, *64*(3), 138–146.

Cumming, G., & Finch, S. (2001). A Primer on the Understanding, Use, and Calculation of Confidence Intervals that are Based on Central and Noncentral Distributions. *Educational and Psychological Measurement*, *61*(4), 532–574.

Gasser, T. M. (2012). *Rechtsfolgen zunehmender Fahrzeugautomatisierung: Gemeinsamer Schlussbericht der Projektgruppe*. *Berichte der Bundesanstalt für Straßenwesen - Fahrzeugtechnik (F): Vol. 83*. Bremerhaven: Wirtschaftsverlag NW.

Gold, C., & Bengler, K. (2014). Taking Over Control from Highly Automated Vehicles. In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.): *Vol. [8]. Advances in human factors and ergonomics 2014, Advances in human aspects of transportation, Part II* (pp. 64–69). Louisville, KY: AHFE Conference.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942.

Kline, R. B. (2013). *Beyond significance testing: Statistics reform in the behavioral sciences* (2nd ed.). Washington, D.C: American Psychological Association.

Körber, M., & Bengler, K. (2014). Potential Individual Differences Regarding Automation Effects in Automated Driving. In C. S. G. González, C. C. Ordóñez, & H. Fardoun (Eds.), *Interacción 2014: Proceedings of the XV International Conference on Human Computer Interaction* (pp. 152–158). New York, NY, USA: ACM.

Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The Influence Of Age On The Take-Over Of Vehicle Control In Highly Automated Driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *39*, 19–32.

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, *6*(3), 299–312.

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2. ed.). San Diego, CA: Academic Press.

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The Time Has Come: Bayesian Methods for Data Analysis in the Organizational Sciences. *Organizational Research Methods*, *15*(4), 722–752.

Kruschke, J. K., & Liddell, T. M. (2015). The Bayesian New Statistics: Two Historical Trends Converge. *SSRN Electronic Journal.*

Kyriakidis, M., Happee, R., & Winter, J. de. (2015). Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transportation Research Part F: Traffic Psychology and Behaviour*, *32*, 127–140.

Merat, N., Jamson, A. H., Lai, F., & Carsten, O. (2012). Highly Automated Driving, Secondary Task Performance, and Driver State. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *54*(5), 762–771.

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2015). The fallacy of placing confidence in confidence intervals. *Psychonomic bulletin & review*, *23*(1), 103–123.

National Highway Traffic Safety Administration. (2013). *Preliminary Statement of Policy Concerning Automated Vehicles*. Washington, DC. Retrieved from www.nhtsa.gov/staticfiles/rulemaking/pdf/Automated_Vehicles_Policy.pdf

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In Kurt Hornik, Friedrich Leisch, & Achim Zeileis (Eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003).* Vienna, Austria.

Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How Traffic Situations and Non-Driving Related Tasks Affect the Take-Over Quality in Highly Automated Driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 2063–2067.

Reichardt, C. S., & Gollob, H. F. (1997). When confidence intervals should be used instead of statistical tests, and vice versa. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 259–284). Mahwah, NJ: Erlbaum.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2010). The impact of systematic variation of cognitive demand on drivers' visual attention across multiple age groups. In *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting 2010* (Vol. 54, pp. 2052–2055).

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, *16*(2), 225–237.

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2015). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods.*

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221–257). Mahwah, NJ: Erlbaum.

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221.