Modeling dual-task performance: *do individualized models predict dual-task performance better than average models?*

Wenjin Cao

---

dr. Chris Janssen
dr. Ben Rin

1 February 2017

Bachelor Artificial Intelligence, University of Utrecht
7,5 ECTS

Table of contents

1 Abstract

Understanding multitasking can be a complicated venture. The goal of this paper is to see whether using individual parameters for modeling dual-task will lead to better predictions of individual performance compared to using the global total average of all participants. It is expected that modeling individual skill will lead to more accurate models because individual parameters will lead to closer fits between individual data points and their models. Therefore, using individual parameters might provide a better explanation of the adaption of different strategies among participants. Data from a study involving 12 participants performing a phone-driving task has been used. The model consists of a driving and dialing model. The results of individualized performance show that using individualized parameters don't necessarily provide more accurate model predictions than using global parameters. This implies that factoring individual skill might not be very useful when modeling dual-tasking performance, however it does tell us an interesting story about the whether there are other individual factors we should take into account. The implication is that it's still too simplistic to look at just average performance to explain multitasking behavior, so it could be interesting to take a closer look at other individual factors in future research.

2 Keywords

Multitasking, driving, cognitive modeling, multitasking strategies, individual differences.

3 Introduction

Multitasking comes as natural as breathing. At this very moment, you might already be doing so; perhaps taking a look at what's going on outside, or taking a sip from your cup of coffee. This all takes place without too much effort. Multitasking can however get quite complicated, as it involves a wide variety of aspects and complex interactions. For example, when you are riding your bike and using Whatsapp to discuss what to cook tonight. You're performing two tasks at the same time that require a lot of skill and attention. You need to be able to observe responses, move your fingers over the screen, type letters, somehow keep track of the road, and make sure you don't lose balance etc. This can possibly lead to quite dangerous situations in traffic. So there's a lot of research on multitasking behavior during the use of phones while driving. Multitasking while driving and drivers distraction has attracted attention on both a national and international level (SWOV 2016, WHO 2011). Research has shown that driving while performing phone visual manual sub-tasks such as texting or dialing is related to significantly increased risk for accidents, possibly more so than talking and listening during driving (Hickman & Hanowski 2012, Fitch et all 2013, Tivesten & Dozza 2014).

There are varying theories on the way multitasking works and how to differentiate strategies. Multitasking can basically be viewed as a management of attention or switching between multiple activities. One perspective from which we can understand multitasking, is through the theory of threaded cognition. This basically means that while multitasking, we have multiple independent threads of thought running at the same time. While simultaneous, this does not always mean literal parallel processing among threads. One has to view it as an interleaving between multiple threads associated with tasks. (Salvucci & Taatgen 2008).

One way to consider multitasking, is by differentiating between two different multitasking behaviors along a time and space continuum, namely concurrent and sequential at both ends. We say concurrent when people are for example driving and talking, where they have to switch between tasks in manners of (sub-)seconds. And we say sequential when people switch between tasks after longer periods of time, with the occurrence of possibly overlap (Salvucci & Taatgen 2011).

Another way is saying that there are 3 different strategies, namely sequential, parallel and interleaving (Adler & Benbunan-Fich 2012). We can consider the task switching paradigm, which was devised after observing that when interleaving between tasks, this leads to decrease in performance of the tasks (Jersild 1927). The alternating-task paradigm is a continuation of this paradigm, and considers the fact that there is an 'additional 'switch cost' required when interleaving between tasks (Rogers & Monsell 1995).

From this point, it would be interesting to see what influences the way people multitask during driving and dialing on their phones. Research has shown that multitasking strategies during these type of tasks is influenced by cognitive and motor chunk boundaries or so called 'breakpoints'. It has been shown that this so called switch cost can be minimized when interleaving occurs at such 'breakpoints'. There's an tendency to interleave after completing sub-tasks (Janssen, Brumby & Garnett 2012). Furthermore, task priority plays a big role as well; for example whether participants have to prioritize driving or dialing to meet certain performance objectives in the dialing and driving dual-task (Janssen & Brumby 2010). Most importantly, research on dialing and driving has shown that phone numbers which have been structured to reflect cognitive and motor cues influence interleaving strategies (Janssen, Brumby & Garnett 2012).

Chapter 3 of Janssen (2012) and research from Janssen, Brumby & Garnett (2012) has expressed the desire to model individual skill as a alternative way to explore dual-task strategies, since there can be large individual differences in dual-task performance, which could be anticipated using single-task parameters (Watson & Strayer, 2010). It can be quite a challenge to approach individual behavior using modeling. It is possible that participants adjust their strategies to individual skill, leading to the importance of explaining performance through individual cognitive, perceptual and motor skills. Specific research has shown that the usual way to model average performance of participants does not lead to the most accurate model. Modeling individual skill seems to lead to better fits between models and participant data. This way, specific strategies will become more well represented (Zhang & Hornof 2014).

This thesis will be primarily based on research by Janssen, Brumby & Garnett (2012), which explores the influence of priorities, cognitive and motor cues during dual-task. Their dual-task consists of a dialing and driving task. In this case, it might be interesting to have a look at how individual characteristics such as typing speed affect performance during dual-tasking. It's expected that performance of participants during preluding single-tasks might give a good pointer of performance during dual-tasking, and explain differences in dual-tasking between participants.

So what we want is to find out if modeling individual skill such as typing speed will provide a better way of explaining performance. Will this be reflected when dual-tasking models are adjusted for individual typing speed? It would be interesting to check when we adjust for each individual speed parameter, will this lead to any difference for predicted optima in comparison to human data? One prediction is that individual models will fit closer to their corresponding human average compared to the global average model. We might observe that participants with quicker typing speeds will have better performance on the driving task compared to slower participants since they are probably interleaving less between tasks.

The study of Artificial Intelligence aims to understand the way humans reason, think, perceive, and use language amongst others and implement this into programs which simulate such behavior. So from an Artificial Intelligence perspective, performing this kind of research can prove to be valuable in our quest to to understand and model human cognition. We want to know how human behavior can be captured by cognitive models, and therefor devise better theories on human cognition. If we know more about the way humans multi-task, perhaps we can think ways of improving this, or anticipating human mistakes using computer programs during important events such as car driving. And if we assume there is not really such a thing as a real 'average user', perhaps we require programs to be adaptive to individual users instead. You can think of many possibilities such as the design of interfaces or generating special levels in computer games.

This leads to our final main question: will using individual parameters, specifically typing speed, for modeling dual-task lead to better predictions of individual performance compared to using the global parameters of all participants?

4 Experiment

In this paper we will model empirical data of 11 participants from Janssen et all (2012) and briefly summarize the method and results they've used. The research was set up with a 2 x 2 (Phone Number x Task priority) within-subjects design. Task priority was either safer driving or faster dialing, over a congruent and incongruent phone number (07333-888111 and 07722-229944) using a Nokia 6300 phone. Janssen et all considers the effects of chunk boundaries in a phone number, and distinguishes between so called cognitive cues motor cues. The numbers used in this experiment contain one chunk boundary, namely at 3-8 and 2-2. We have a cognitive cue when having to switch between the numbers 2-2 at the chunk boundary of the incongruent phone number 07722-229944. We have a motor cue at 3-8 in the congruent phone number 07333-888111, since it requires a new motor action, since the participant has relocate it's finger to press a different number .

Steering performance was assessed through mean absolute lateral deviation of the simulated vehicle. Participants had to do 10 single-task steering trials first, and two blocks of experimental trials. There were dialing practice trials, single-task steering trials, single-task dialing trials and dual-task steering-and-dialing trials.

Participants practiced to memorize the phone numbers by typing the first part before the chunk boundary (having the rest invisible), and then typing the second one (having the first part before the chunk invisible). In the single-task dialing trials they had to dial the phone number as fast as possible. Data from the single-task trials will be used to determine inter-key press press intervals; the interval between key presses.

Performance was measured from the moment a participant pressed the first digit, until the last digit of a phone number. Trials which contain typing errors have been excluded as well. Main effect of task priority was that dialing times were faster when dialing was prioritized. Overall performance gives us as dialtime for the dialing focus M = 5.0 s, SD = 0.8 s. Dialtime for the steering focus M = 7.6 s, SD = 1.6 s. Lateral deviation for the dialing focus M = 0.77 m, SD = 0.23 m. And lateral deviation for the steering focus (M = 0.49 m, SD = 0.12 m). Steering was better when this was prioritized. Participants made more active steering movements during the steering priority. There were no significant main effects of phone number on both tasks. The participants seemed to adjust their strategies depending on strategies.

An significant main effect was found among task priority, phone number, and digit type on the duration of inter-key press press intervals. Higher times were found for the first number and chunk boundary in dual-task, this was also observed in the single-task.

The phone numbers can be categorized in five different digit types, namely a chunk boundary, first of repetition (in a series), repeating digit, the very first digit, and a second digit. For clarity, it is shown below how it is categorized. For the phone number 07333-888111:

0 = very first digit
7 = second digit
3 = first of repetition
3 = repeating digit
3 = repeating digit
8 = chunk boundary
8 = repeating digit
8 = repeating digit
1 = first of repetition
1 = repeating digit
1 = repeating digit

The differentiation between digit types seemed to be an important factor in performance. Inter-key press press time at the chunk boundary was higher in when dialing the congruent number compared to the incongruent number. Participants were found to make more steering movements were elevated inter-key press press times were found, especially at the natural break-point, such as the chunk boundary or when there was a number change. More steering movements also occurred at the chunk boundary when dialing when dialing the congruent number compared to the incongruent number.

5 Model of Average performance

5.1 The model

As a starting point, the model of average performance used by Janssen, Brumby & Garnett (2012) was also used for this paper, based on the framework of cognitively bounded rational analysis, meaning this approach tries to specify predictions through the analysis of payoff achieved by alternative strategies (Howes, Lewis & Vera 2009).

The is a dual-task model consists of a steering and a dialing model. The steering model performs steering updates to try and keep the car back to center, and the dialing model takes the amount of time needed to press a single key, the key press time, as a parameter. The model has a central processing bottle neck, meaning that both tasks can't be performed at the same time.

5.2 Dialing model

The dialing model uses average participant inter-key press times as calculated from the single-task condition. These values denote the time it takes to press each key type. It was assumed that switching attention between the road and phone takes 200 ms, and that the time needed to resume typing at a chunk boundary takes 100 ms, this retrieval cost was subtracted from the first digit and chunk boundary.

Since each phone number has 11 digits, and it is possible to either dial or drive after each digit, there are in total 2^10 = 1024 unique possibilities of interleaving strategies. The original model uses 12 different possible numbers of steering updates, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, which gives a total of 12,227 strategy alternatives.

The amount of time needed for each steering update was set from 250 to 3000 ms with increments of 250 ms. To get a reliable estimate of average performance, the number of simulations was set at 50 per participant per strategy alternative.

Average performance was reported for each strategy alternative for the participant mean per phone number.

5.3 Driving model

The parameters of the steering model consists of a updating value of 50 ms, combined with a new value taken from a Gaussian distribution using a 0.00 m/s mean and 0.13 m/s standard deviation. Lateral velocity is adjusted every 250 ms, which a maximum at 1.7 m/s. We take a set of 1, 2, 4, 6, 8, 10, and 12 as steering update options. 12 was found to be the upper limit for discernible lane keeping performance (Brumby & Salvucci & Howes 2009). These represent the amount of successive steering updates after each 250 s. A value taken from a Gaussion distribution using a 0.00 m/s mean and 0.10 m/s standard deviation was used to model human variability.

Velocity = 0.2617 x LD2 + 0.0233 x LD - 0.022

Average lateral velocity during steering has been modeled according to an equation developed by Brumby, Salvucci & Howes (2006) using previous experiments (Salvucci & Macuga, 2002). The model predicts that when lateral deviation from the center increases, lateral velocity also increases in order to bring the car back to the center.

5.4 Result of model of Average performance

While human mean performance in the research of Janssen, Brumby & Garnett (2012) shows that the human mean data lies just on the outside of the trade-off curve (where optimum performance is predicted) of the model of average performance. Each point in the plot reflects an unique strategy. Note that the original model as discussed in Janssen, Brumby & Garnett (2012) uses a full set of 12 steering updates, however we plot it over the model of average performance using 7 steering updates, since this was also used for the model of individual performance, as shown below in figure 1 and 2.

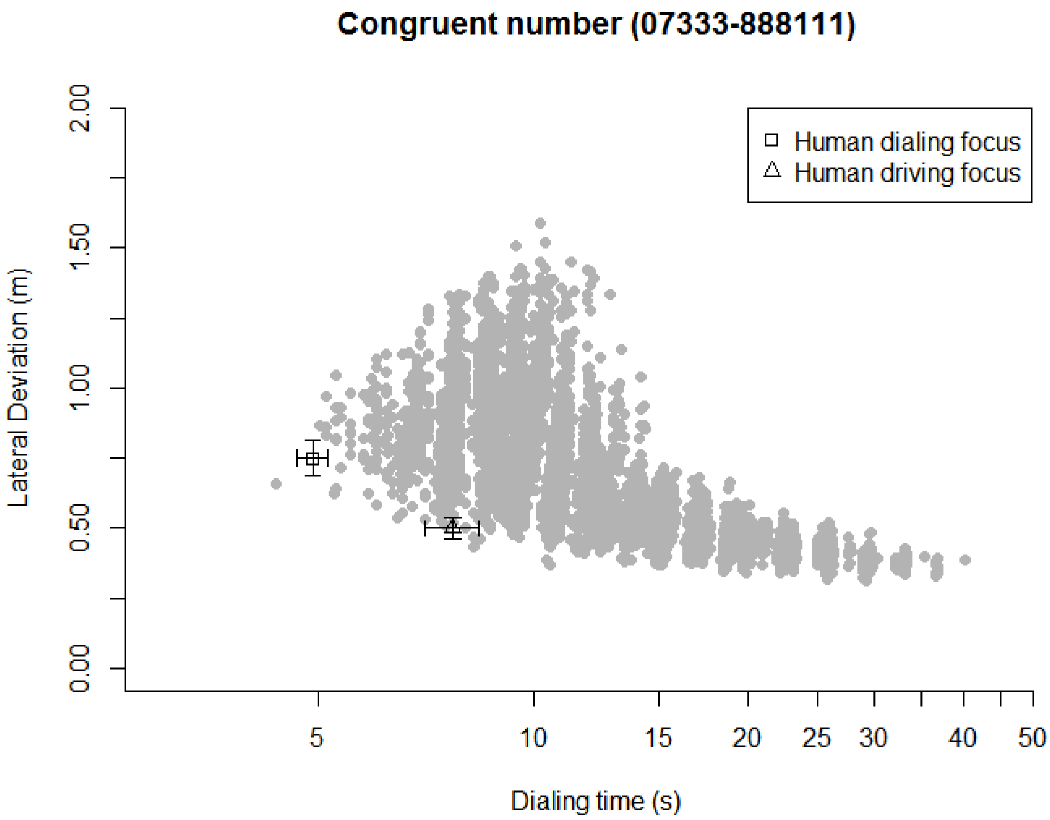


Figure 1. *Congruent model, 50 simulations. Human dialing focus. Error bars represent standardized error of the total participant mean.*

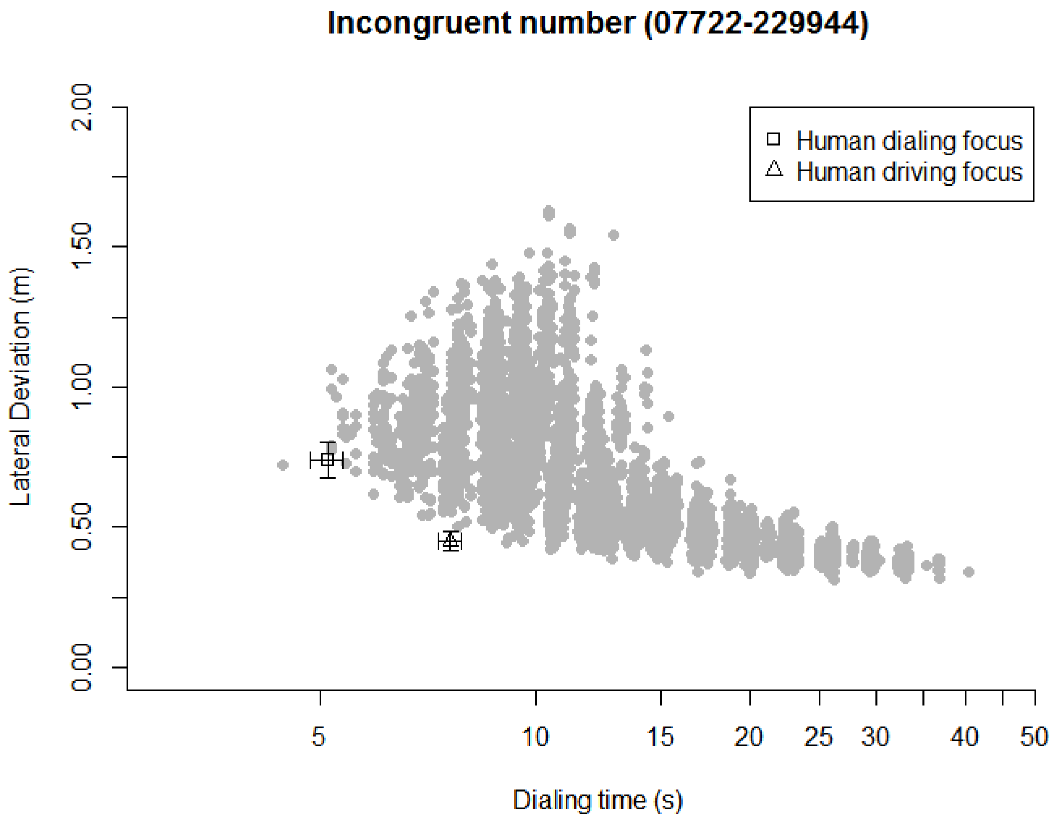


Figure 2. *Congruent model, 50 simulations. Human dialing focus. Error bars represent standardized error of the total participant mean.*

We now ask ourselves the question, how does he individual participant data fall within the average model performance? The average dial times and lateral deviations of the individual participants have been calculated by taking data from the dual-task and excluding trials which contained dialing errors, as has also been done for the average mean of all participants in Janssen, Brumby & Garnett (2012).

We notice for the congruent number- dialing focus in figure 3, participants 1, 4, 5, 6, 7, 8, 9 and 12 lay outside the performance cloud and trade-off curve. However we see that for the same participants in the driving focus in figure 4, only participants 4 and 8 are really outside the performance cloud and trade-off curve. So there are more human data points off in the dialing focus condition than driving focus.

For the incongruent number- dialing focus in figure 5, we see that participants 1, 4, 5, 7, 8, and 12 lay outside the performance cloud and trade-off curve. We see that for the same participants in the driving focus in figure 5, participants 1, 2, 4, 7, 8 and 12 are outside the performance cloud and curve.

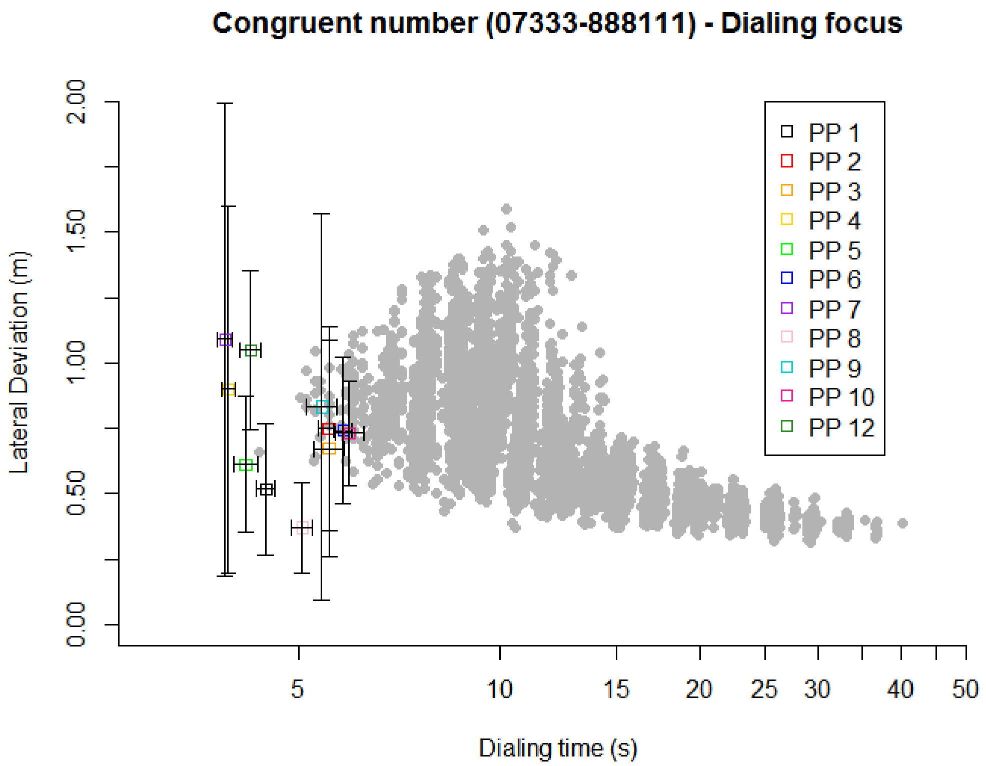


Figure 3. *Congruent model, 50 simulations. Human dialing focus. Error bars represent standard deviation of individual means.*

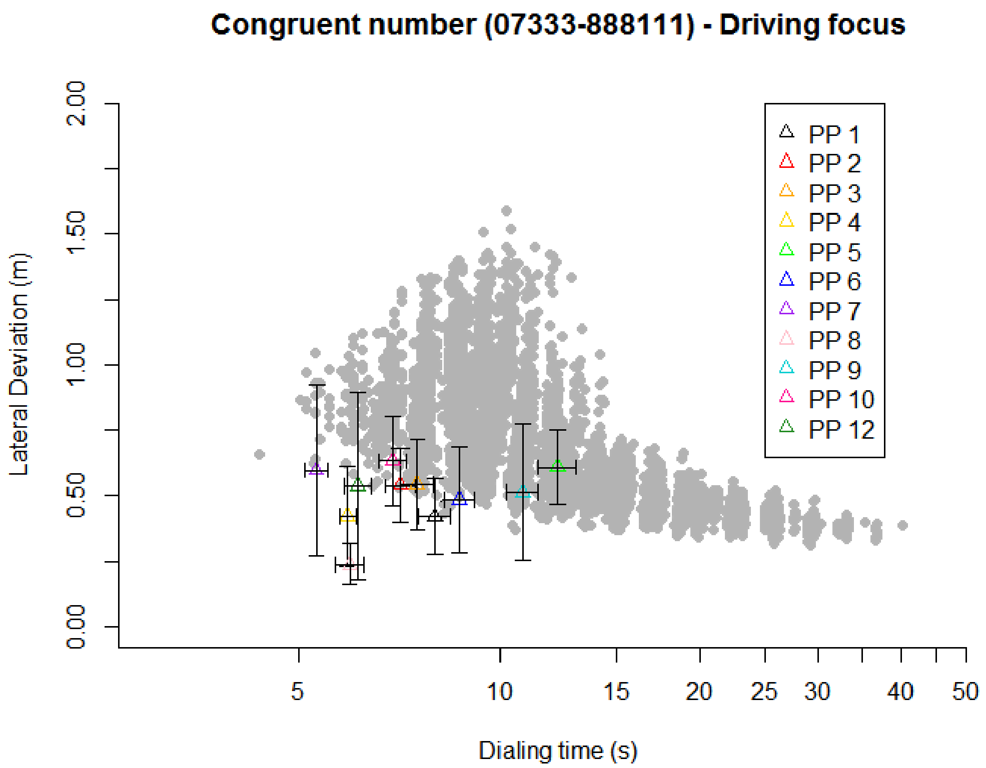


Figure 4. *Congruent model, 50 simulations. Human driving focus. Error bars represent standard deviation of individual means.*

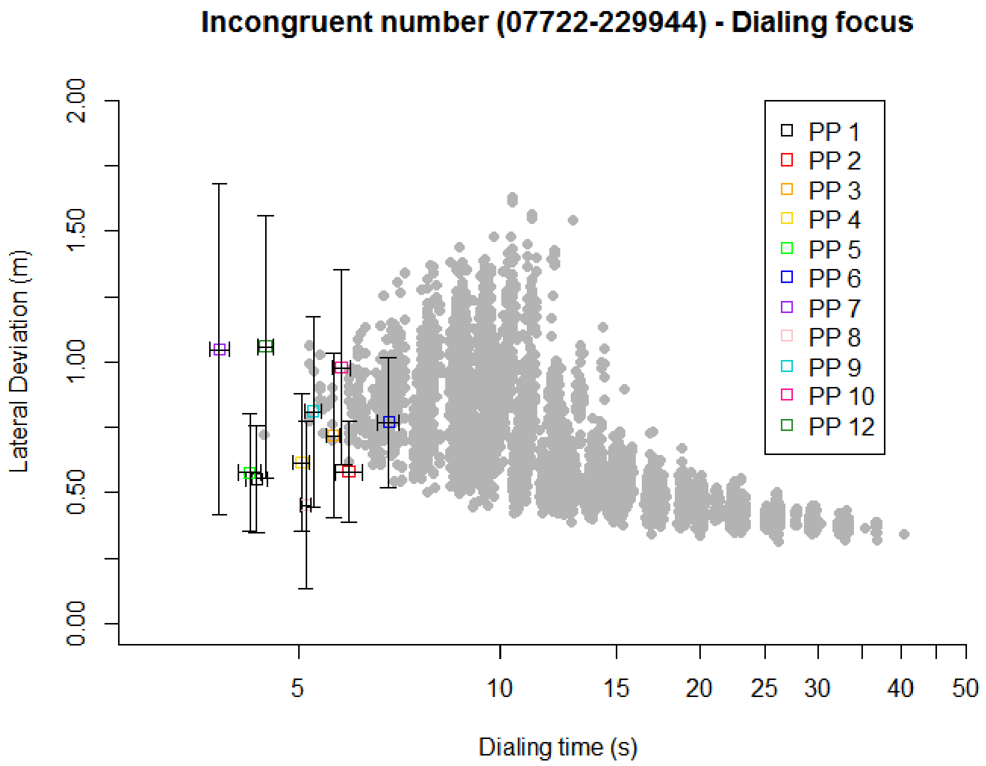


Figure 5. *Incongruent model, 50 simulations. Human dialing focus. Error bars represent standard deviation of individual means.*

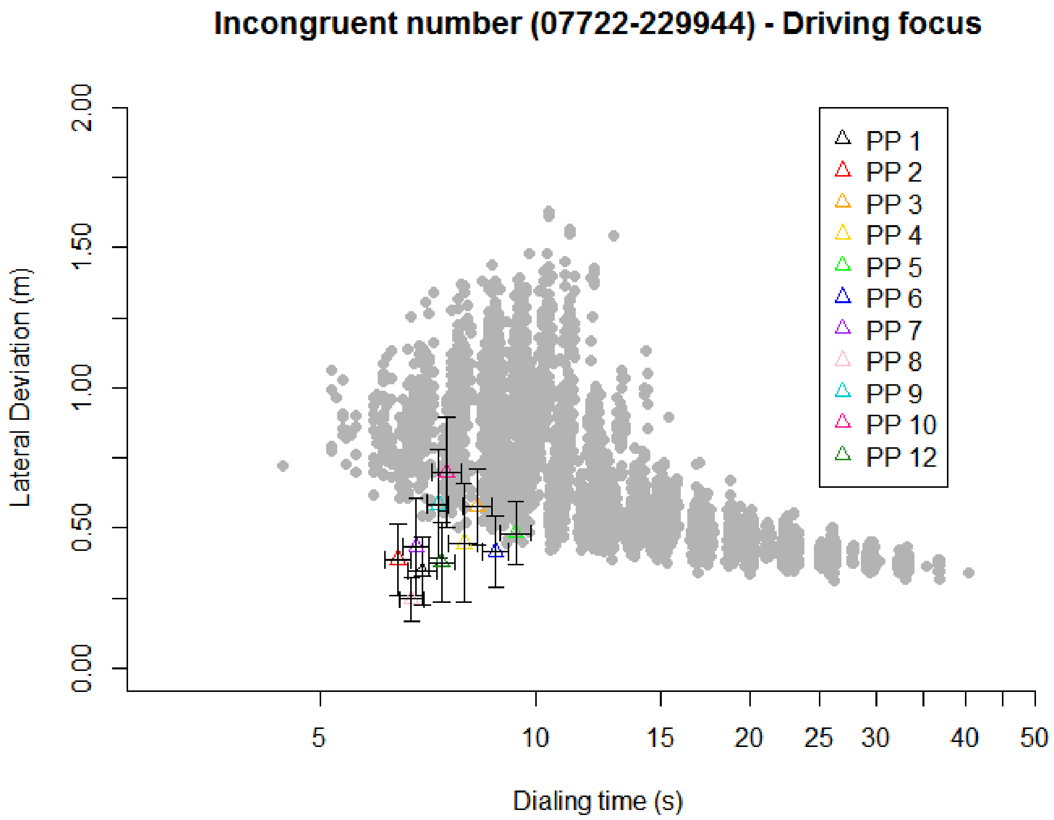


Figure 6. *Incongruent model, 50 simulations. Human driving focus. Error bars represent standard deviation of individual means.*

Looking at all of the human data and comparing it with both the regular and individualized models, we can say that when we have 11 participants, with 2 different phone numbers (congruent and incongruent) and 2 different priority conditions (dialing and driving focus), we get 11x2x2 = 44 total situations.

Of these situations or specifically human participant points, so we can summarize by saying for the congruent phone number, there are in total 7 points that are still within the standard deviation of the predicted strategies in the model. There are 5 points that fall far outside the spread of the cloud. And there are 10 points where they lie inside the model.

And for the congruent phone number, we can summarize there are in total 10 points that are within the standard deviation of the predicted strategies in the model. There are 4 points that fall far outside the spread of the model. And there are 8 points where they lie inside the model.

So we see that that the average performance model does not fully account for individual performance. We see a lot of points lie far outside the points as predicted by the average model. We can conclude that the model fits the average participant mean, but does not fit the individual participant means. This is why we want to use individualized parameters to create more accurate models that might provide a better fit with individual data.

6 Model of Individual performance

6.1 The individual model

What we would like to know now, is what we get when we use individual inter-key press times as parameters instead of the global parameters, as used by the model of average performance. So for digit types, we extracted from the 'phonepretest' condition: chunk boundary, first of repeating series, repeating digit, very first digit, and second digit. These parameters have been calculated by looking at each individual, taking the average of each trial of the individual (trials with dialing errors have been excluded), and then taking the total average of all trials.

These were imposed on the original model. We do this because we want to check whether using individual parameters from the single-task will lead to a closer fit between the individual human dual-task performance, and the corresponding individual model which models dual-task performance.

The individual parameters are shown in the appendix 3 and 4. We see that our slowest participant (highest average key press time) for the congruent condition is participant 10, and the fastest (lowest average key press time) is participant 12. Our slowest participant for the incongruent condition is participant 12, and the fastest is participant 8.

Besides using individual parameters, it was also decided to use a subset of the steering update options instead of the complete one used by Janssen, Brumby & Garnett (2012), since the full set led to a undesirably long simulation time of about 20 hours per individual, and the subset averages at about 8 hours per simulation. Using a smaller set of parameters won't lead to a significant change in the arrangement of the points in the model, it will only be slightly less detailed since there are less strategies to go through.

So since each phone number has 11 digits, and it is possible to either dial or drive after each digit, there are in total 2^10 = 1024 unique possibilities of interleaving strategies. Since we used a subset of the steering update options, we have a set of 7 different possible numbers of steering updates, namely 1, 2, 4, 6, 8, 10, 12- 7 different alternatives, giving us 7189 strategy alternatives.

Average performance was reported for each strategy alternative per participant per phone number.

6.2 Results model of Individual performance

These figures show the performance space for strategy alternatives of the congruent and incongruent phone numbers.

See appendix 1 and 2 for full overview of all plots.

We first want to take a look at several participants to see what kind of result modeling for their individual parameters has given us compared to the model of average performance. We look at some participants who originally had its individual mean outside and inside the predicted strategy space of the average model. We take a look at one of the participants, participant 4, because this one has previously shown to have both the dialing and driving focus data outside the clouds and trade-off curve of the model of average performance. We see the same again when modeling with individualized parameters, as shown by figure 5 .

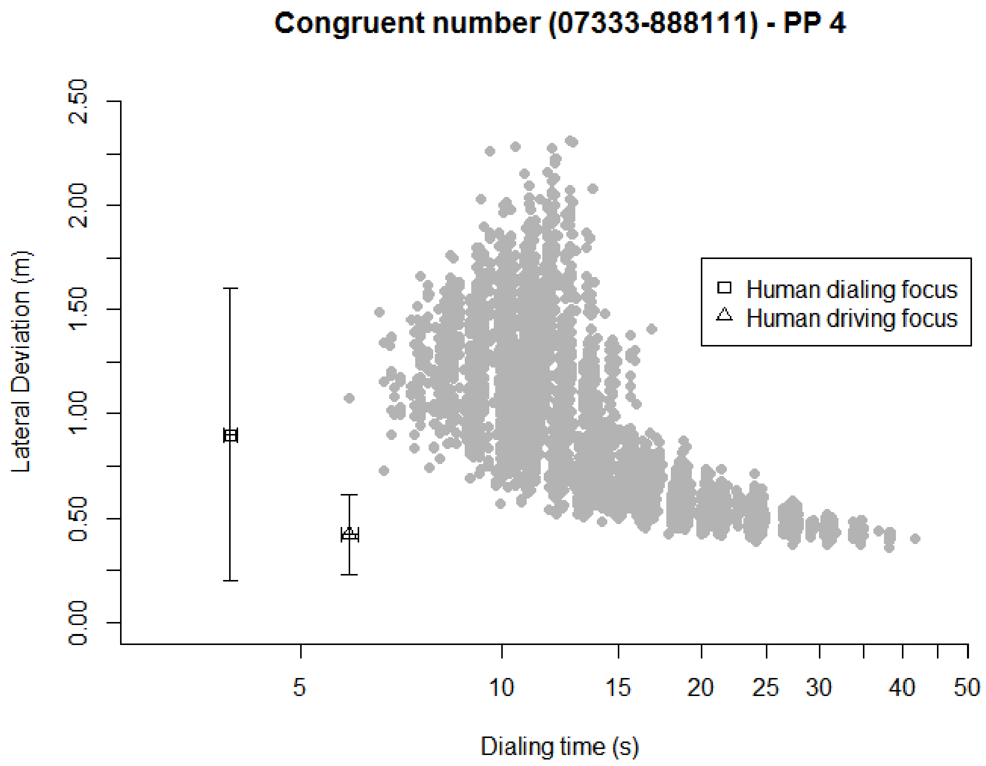


Figure 7. *Congruent model, 50 simulations, participant 4. Error bars represent standard deviation of individual mean.*

We also want to take a look at a participant who had their data well inside the predicted strategies of the average model, and see how this has changed for its individual model. For participant 10, we observe that it's performance was well inside the predicted strategies of the average model, however the individualized model is far off from it. Oddly, this is not the case for the incongruent number condition for participant 10, where it's shown to have moved outside.

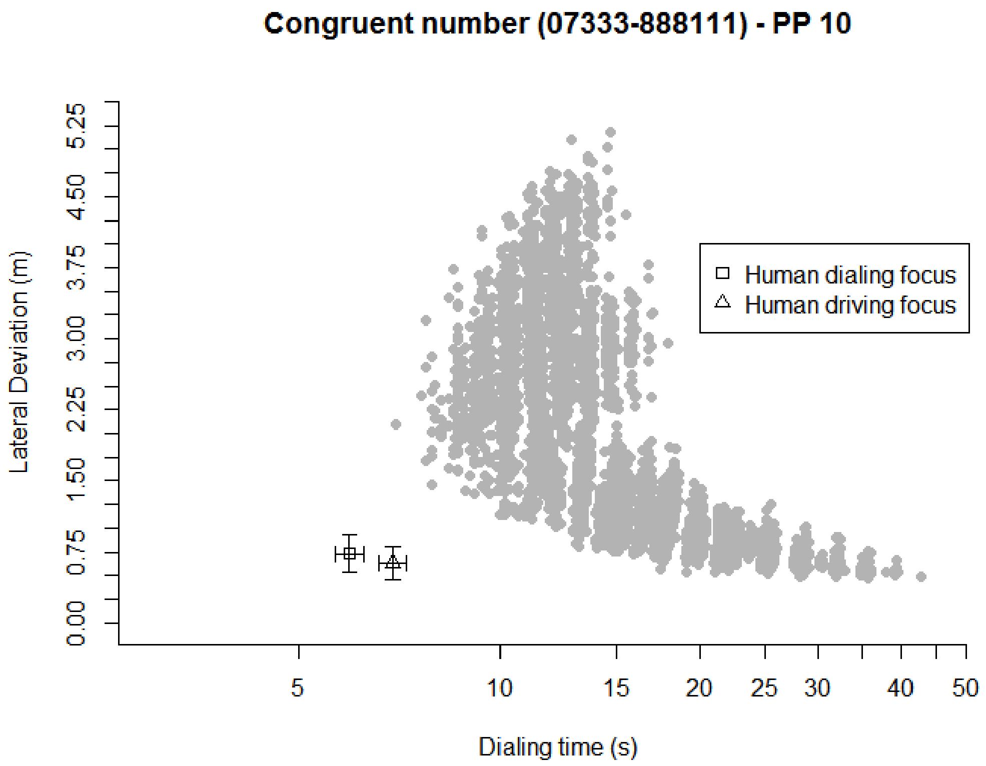


Figure 8. *Congruent model, 50 simulations, participant 10.*

For one of the participants who had its data point for the congruent number- dialing focus of relatively far off from the cloud, we notice for participant 12 that when it's human mean has moved closer to the predicted strategies of the individual m odel. And it is now within the standard deviation for the human dialing condition.

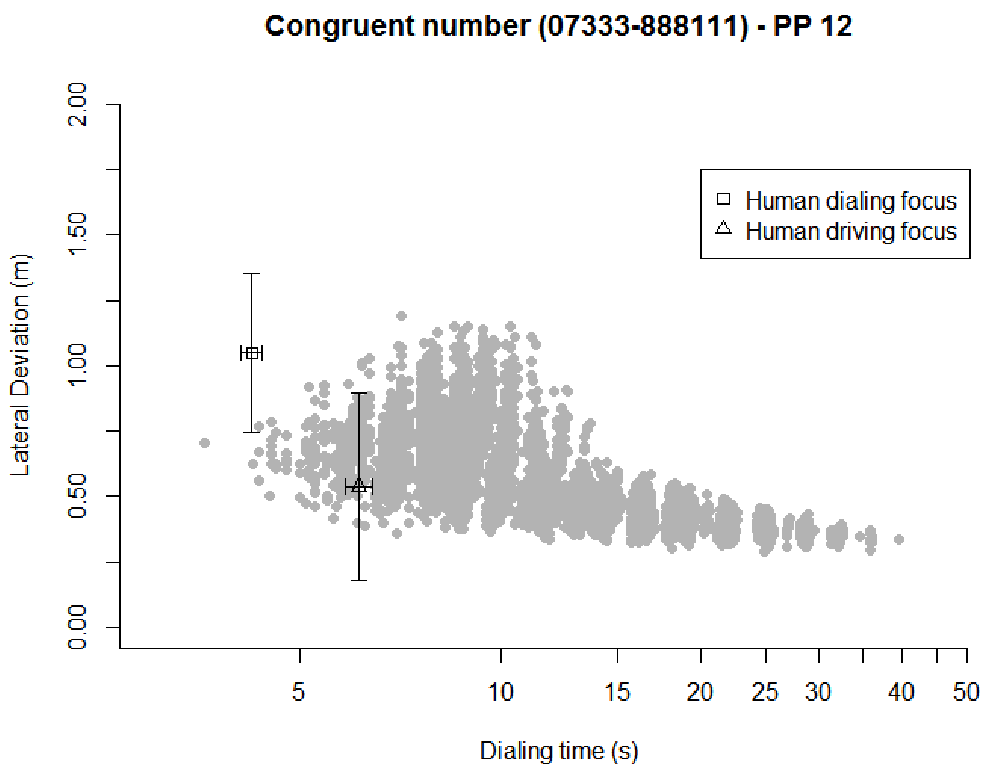


Figure 9. *Congruent model, 50 simulations, participant 12.*

We check participant 1 for the incongruent number, which had the points for both dialing and driving focus outside the cloud. We notice this is not the case with the individualized model as shown in figure 9. The individualized model seems to provide a better fit.

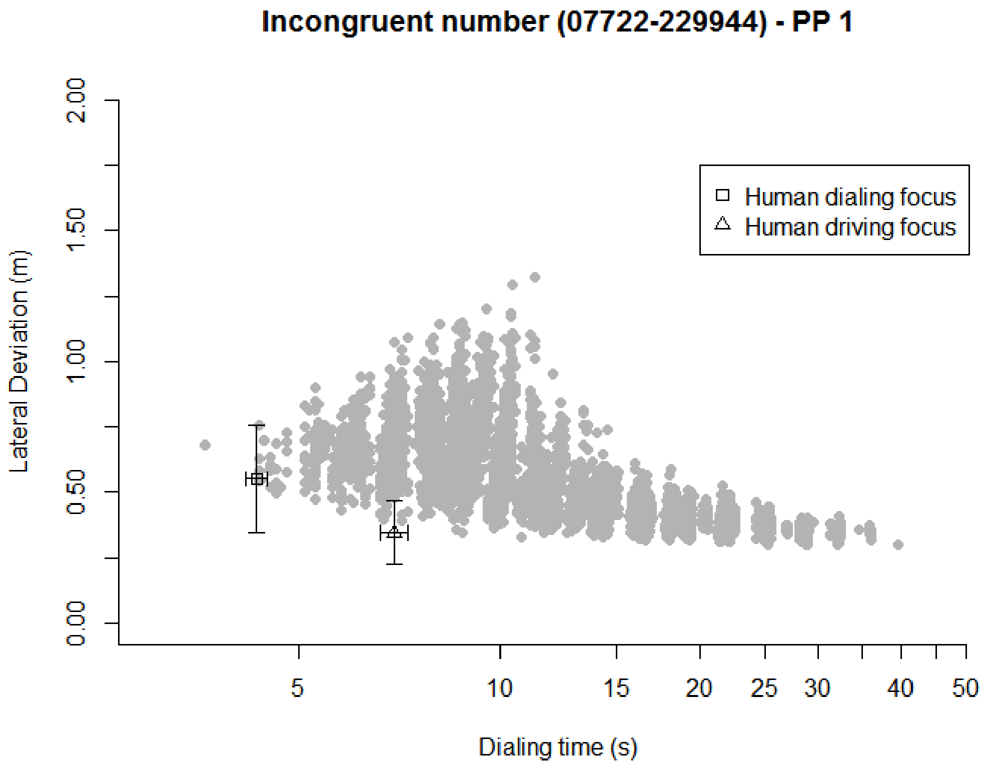


Figure 10. *Incongruent model, 50 simulations, participant 1.*

And similarly, for the incongruent number, participants 7 and 12 don't display a closer fit between the human data points and their individualized models.

Looking at all of the human data and comparing it with both the regular and individualized models, we can say that when we have 11 participants, with 2 different phone numbers (congruent and incongruent) and 2 different priority conditions (dialing and driving focus), we get 11x2x2 = 44 total situations.

Of these situations or specifically human participant points, so we can summarize by saying for the congruent phone number, there are in total 6 points that are still within the standard deviation of the predicted strategies in the model. There are 8 points that fall far outside the spread of the cloud. And there are 8 points where they lie inside the model.

And for the congruent phone number, we can summarize there are in total 5 points that are within the standard deviation of the predicted strategies in the model. There are 8 points that fall far outside the spread of the model. And there are 9 points where they lie inside the model.

So we have a total of 16 points that are completely outside the spread of the model, and it seems that modeling for individual parameters has not lead to obvious improvements of fit between the participant means and their respective individual models. Possibly, most participants eventually execute the dual-task better and quicker than can be predicted on the basis of pretest results. It is interesting to note that participant 9 for example, has both points well within it's own performance cloud for the incongruent phone number. This suggest worse than optimal performance according to the model.

Janssen, Brumby & Garnett (2012) have shown that the mean inter-key press intervals among the participants vary significantly between the single-task and dual-task trials. What we want to do now, is find out how inter-key press times of the pretest and dual-task compare. See appendix 5 for full overview. So similarly to their research, the inter-key press times for the dual task driving focus is considerably higher than the dialing focus, and we see that on an individual level as well. For example, there is a summed difference of 955 milliseconds for participant 4, and 224 milliseconds for participant 7. When looking at overall times, we observe that the dual-task setting yields higher key press times than the single-task. Looking at the differences between these times, it might explain why many human data points didn't fall within their predicted models.

When looking at all participants, there does seem to be an association between higher dual-task inter-keypress times and points laying further outside the model. In appendix 6, marked gray, we can find all the participants who have data points which are noticeably more towards the start of horizontal axis and therefor further away from the model on a horizontal scale, compared with other models. We should observe the same for the incongruent phone number, however here the association seems less pronounced, since we don't observe the same effect on 5, unless we factor out the first point in the plot, which is basically the 0-strategy, where no-interleaving occurred at all.

Lastly, we also try to determine the most nearby point based on the horizontal dial time axis, and try to quantify if the individualized models have made any improvements in predicting the human dual-task data. We calculate these points by subtracting the total human mean dial time from all the dial times of the strategies from their respective models. We determine which point has the smallest absolute difference compared to the human mean. We consider the fact that there are less unique dial times than lateral deviations, therefor we pick the one corresponding with the smallest lateral difference compared to the lateral deviation that corresponds with the human data point. These values can be found in the appendix 7.

Considering the fact that we can also calculate the most nearby point based on lateral deviation, we quickly notice that there are indeed more unique lateral deviations as mentioned before, in such a quantity that it might not be fruitful to consider looking further at this data, since the absolute lateral differences between human data and model data are apparently so small.

So eventually, we compare the individualized and average models by looking at the total average of these absolute differences. The average congruent model has a total average of 170 ms absolute difference compared to 190 ms for the individualized congruent models. The average incongruent model has a total average of 82 ms absolute difference compared to 101 ms for the individualized incongruent models. So this leaves both not necessarily comparatively better.

7 Discussion

7.1 Summary

Looking at the amount of points which we have previously determined to be within standard deviation, outside the spread of the model, and inside, we can see that individualized models seem to have decidedly worse fits that the average model. The individual models had less points that were within standard deviation, and more points that ended up outside the cloud. However the individual models did have one more point inside the cloud, but this is hardly significant. When we look at how dual-task and single-task key press times compare, we see that the dual-task times are higher, so this might explain why many human data points didn't fall within their predicted models. We also see that the individual models don't necessarily lead to a better fit when determining the closest dial time point.

That said, the differences between the pretest and dual-task inter-key press times did seem to explain fairly well how we could have gotten to such discrepancies. Apparently, some participants just simply perform better at the dual-task than they previously did in the single-task.

We previously stated that we might observe that participants with quicker typing speeds will have better performance on the driving task compared to slower participants since they are probably interleaving less between tasks. We did see that 12 had on average the smallest key press time, and it's corresponding individual model indeed shows comparatively less spread on the lateral deviation axis than other models. Similar for participant 10, who was the slowest, and had the most spread. This seems less obvious for the incongruent model.

This implies that modeling for individual parameters doesn't necessarily lead to a better fit between individual human mean data and model data. This is in contrast with what has been suggested by previous research on modeling using individualized parameters. However these researches used different settings, one where participants had to perform mathematical operations while driving (Watson & Strayer, 2010), and one with a tracking and classification task on the computer (Zhang & Hornhof, 2014).

7.2 Limitations

Looking at any possible flaws on hindsight, we can consider the the code written to calculate the keypress times. One explanation that comes to mind for possible discrepancies, is the fact that the per subject keytype averages from the single-task have been calculated in a manner that possibly weighs some values a bit more heavily than others. This was done by calculating per trial values per subject first, and averaging over all trials of the single-task. Although this didn't seem to have led to significant differences compared to the mean values taken from the literature, except for the very first digit, possibly due to subject 10. There we see a noticeable outlier for the first keypress, which leads to the very first keypress average to deviate from the average as stated in the literature.

One possible explanation for this, is that calculating the averages didn't factor in the problem that the computer program for the dual-task would allegedly sometimes take a while to buffer, while the time counter has already been set. This has possibly led to a significantly higher first digit

keypress time for subject 10 for example. So there might or might not have been a problem with determining the correct keypress times from the single task in the first place.
However on an individual level this should not have mattered much. One solution that could have tackled this problem was by using the median function instead of mean function in R, but previous efforts have shown to lead to even more inconsistent values.

Another important issue to point out, is that the very first point of the model is the point where no interleaving occurs, and this was the most nearby point for several participants. Factoring this out, and perhaps taking a closer look at which specific strategies correspond with the human strategies, while also looking at the individual steering events might provide us more insight.

It must also be noted that since not a full set of steering options was used, the model was a bit less detailed. This begs the question whether using the full set would've provided a better anticipation of human data since there would've been more points to compare with. However, in a global sense, it would've not made much difference since the spread of all the different strategies would've remained more or less the same. We took a set of 1, 2, 4, 6, 8, 10, and 12 as steering update options, and only omitted the uneven numbers starting from 2.

8 Conclusion

The goal of was to find whether using individual parameters for modeling dual-task will lead to better predictions of individual performance compared to using the global total average of all participants. We've found that modeling individual skill, specifically typing speed, does not lead to more accurate models Therefor, using individual parameters such as typing speed, does not provide a better explanation of the adaption of different strategies among participants.

What has been previously discussed in the *Discussion* section are all things that should be considered for future research. Improving methodology is a definite must. As for implications in the field of Artificial Intelligence research, we can say that we've gained a bit more knowledge on modeling human cognition. We have found that using individual parameters such as typing speed doesn't seem to fully capture human dual-tasking behavior. So typing speed might be irrelevant.

This however does not imply that looking at other individual factors has become insignificant. The fact that we've been unable to achieve a better model actually leads to the question: what *does* influence dual-task performance, if not for typing speed? Since the average model still doesn't fully account for all individual differences. This might open up more possibilities for research into modeling dual-tasking, and eventually human cognition. What we regularly know as the 'average' person might or might not exist at all. And if not, the best way to deal with this, is by employing AI to tackle individual differences and problems.

Reference

Adler, R. F., Benbunan-Fich, R. (2012), Juggling on a high wire: Multitasking effects on performance, International Journal of Human-Computer Studies, Vol. 70, No. 2, 156-168.

Fitch, G. A., Soccolich, S. A., Guo, F., McClafferty, J., Fang, Y., Olson, R. L., Perez, M. A., Hanowski, R. J., Hankey, J. M., & Dingus, T. A. (2013). The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk. (Report No. DOT HS 811 757). Washington, DC: National Highway Traffic Safety Administration.

Hickman, J. S., Hanowski, R. J. (2012). An assessment of commercial motor vehicle driver distraction using naturalistic driving data. Traffic Injury Prev. ;13(6): 566–574.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. Psychological Review, 116, 717–751.

Jersild, A.T. (1927). Mental set and shift. Archives of Psychology (Whole No. 89, 5–82)

Janssen, C. P., & Brumby, D. P. (2010). Strategic adaptation to performance objectives in a dual-task setting. Cognitive Science, 34, 1548–1560.

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

Janssen, C. P. (2012). Understanding Strategic Adaptation in Dual-Task Situations as Cognitively Bounded Rational Behavior.PhD thesis in Human-Computer Interaction. University College London.

Payne, S. J., Duggan, G. B., & Neth, H. (2007). Discretionary task interleaving: Heuristics for time allocation in cognitive foraging. Journal of Experimental Psychology: General, 136, 370–388.

Rogers R.D., Monsell S. (1995). The cost of a predictable switch between simple cognitive tasks. Journal of Experimental Psychology: General, 124, 207–231

Salvucci, D. D., & Macuga, K. L. (2002). Predicting the effects of cellular-phone dialing on driver performance. Cognitive Systems Research, 3, 95–102.

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

Salvucci, D. D., & Taatgen, N. A. (2011). The Multitasking Mind. New York, NY: Oxford University Press.

Tivesten, E., Dozza, M. (2014). Driving context and visual–manual phone tasks influence glance behavior in naturalistic driving. Transportation Research Part F: Traffic Psychology and Behaviour, 26 (Part A(0)) (2014), 258–272

SWOV (2016). Monitor Verkeersveiligheid 2016: Toename verkeersdoden en ernstig verkeersgewonden . Nederland, Stichting Wetenschappelijk Onderzoek Verkeersveiligheid.

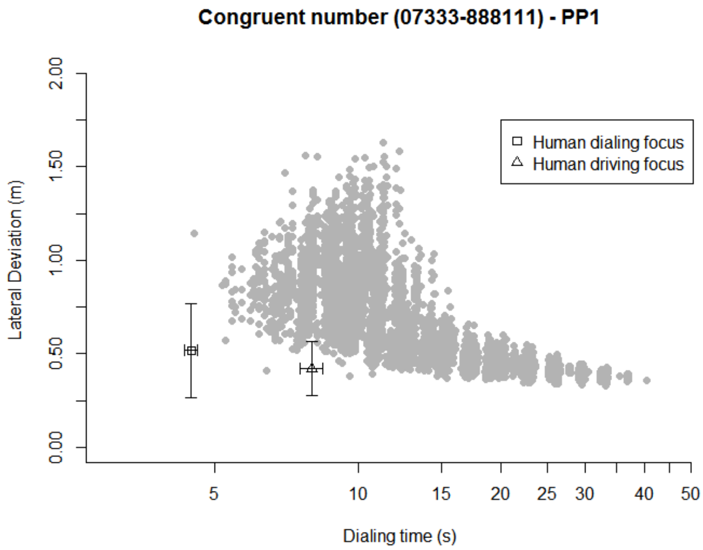
Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. Psychonomic Bulletin & Review, 17, 479–485.

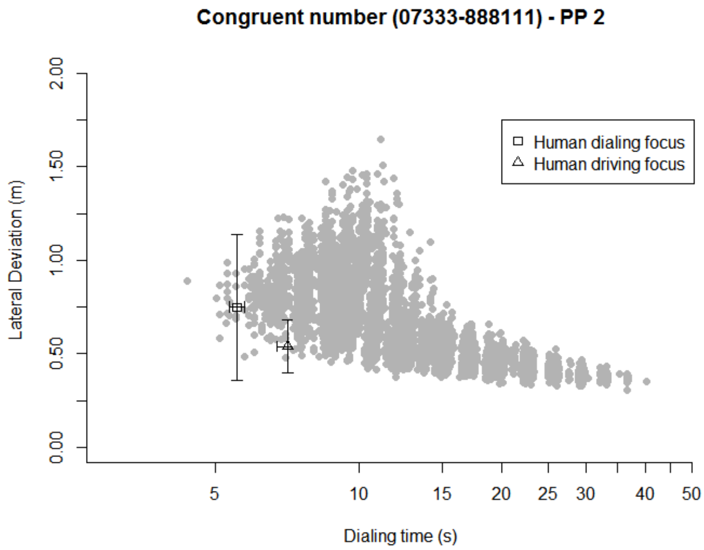
WHO (2011). Mobile phone use: A growing problem of driver distraction. Geneva, Switzerland, World Health Organization.

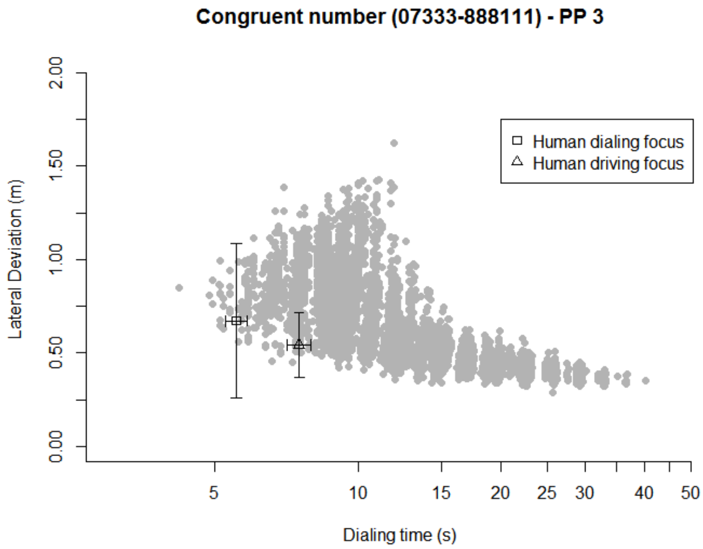
Zhang, Y., & Hornhof, A. J., (2014). Understanding multitasking through parallelized strategy exploration and individualized cognitive modeling. CHI 2014, 3885-3894.

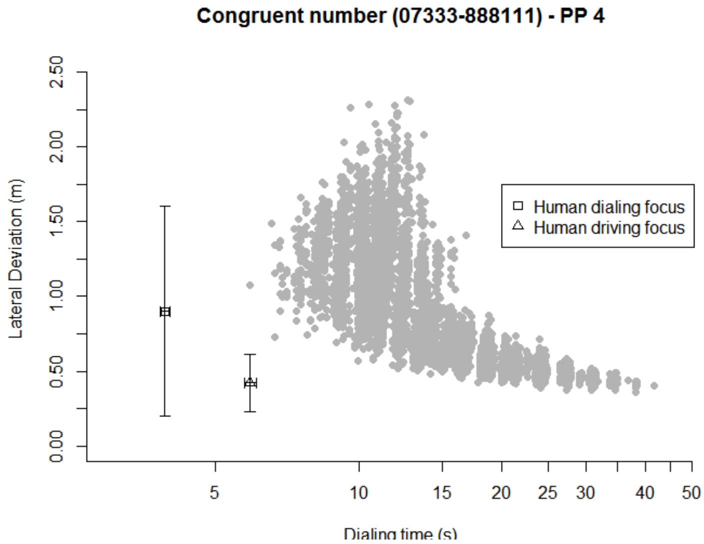
Appendix

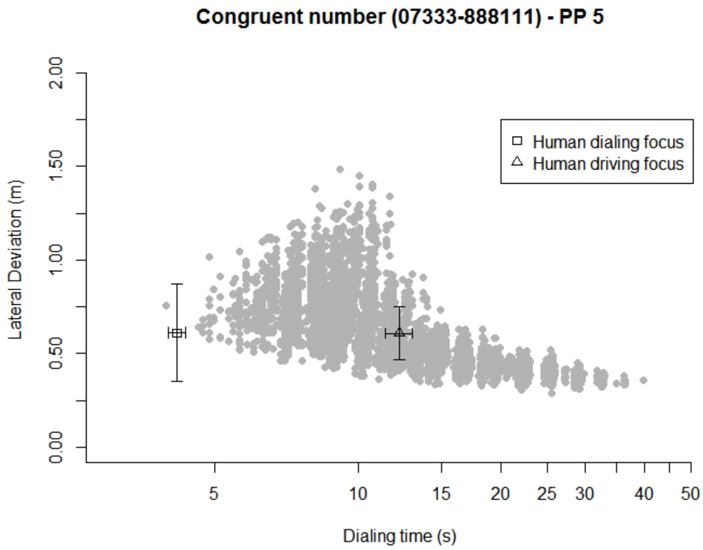
Appendix 1: Congruent individual models. Error bars represent standard deviation.

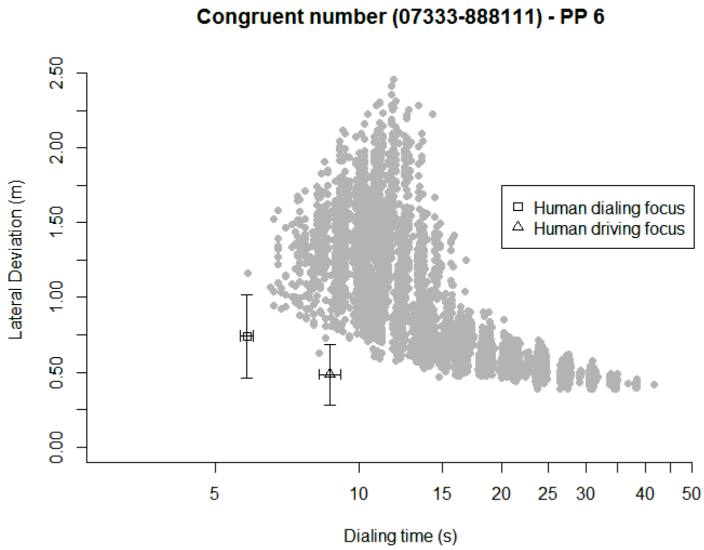


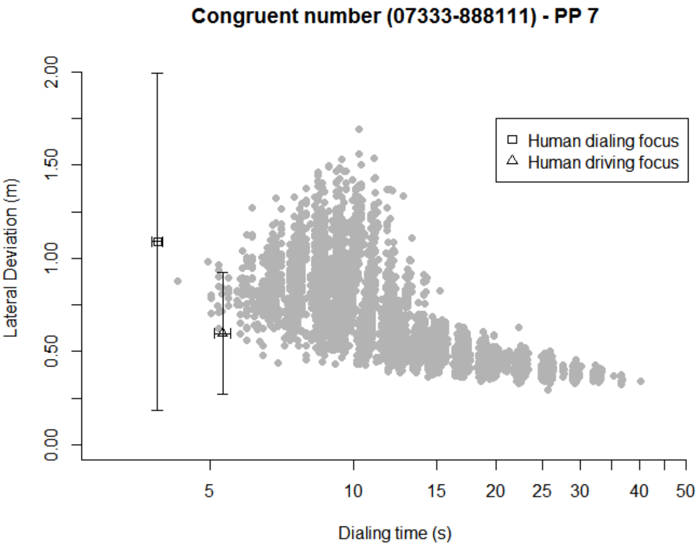


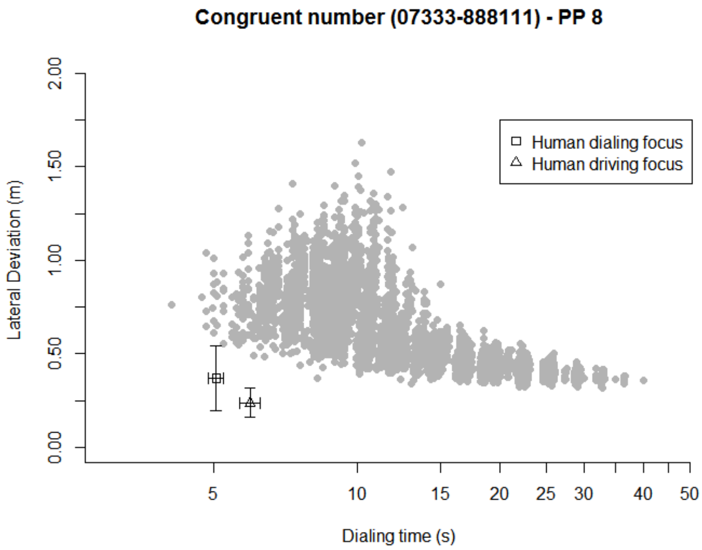


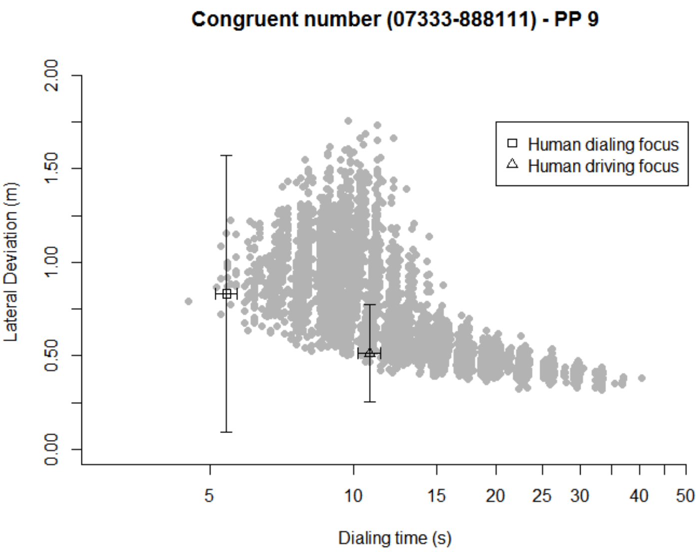


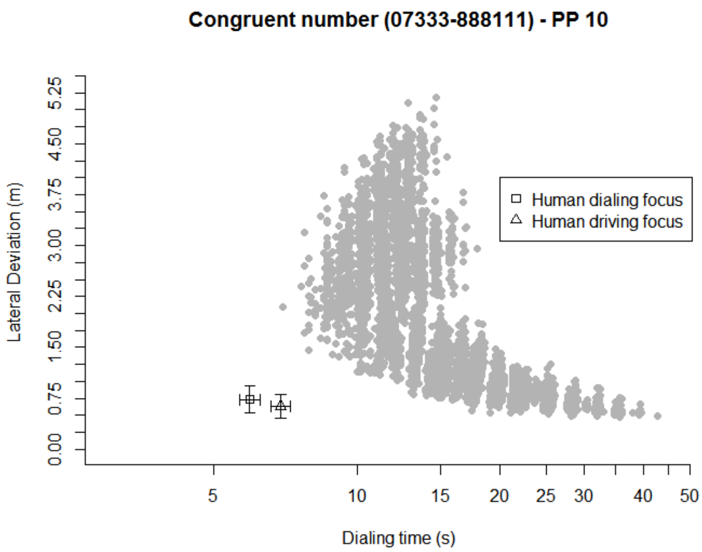


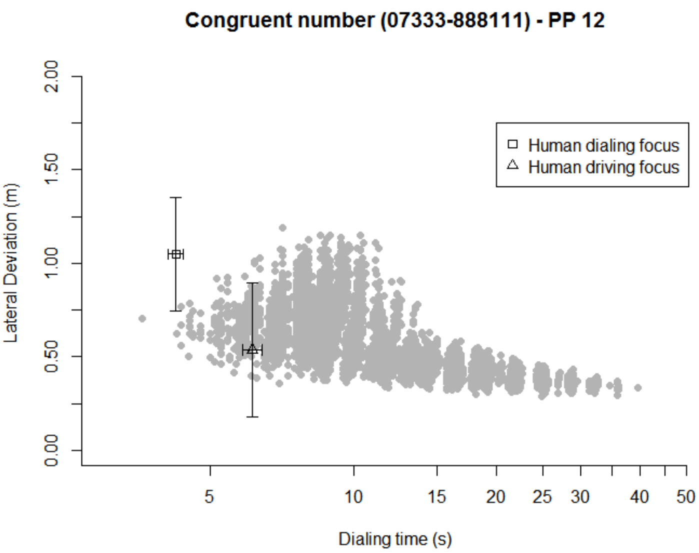



Congruent number (07333-888111) - PP 7
Lateral Deviation (m)
0.00
0.50
1.00
1.50
2.00
Human dialing focus
Human driving focus
5
10
15
20
25
30
40
50
Dialing time (s)


Congruent number (07333-888111) - PP 8
Lateral Deviation (m)
0.00
0.50
1.00
1.50
2.00
Human dialing focus
Human driving focus
5
10
15
20
25
30
40
50
Dialing time (s)

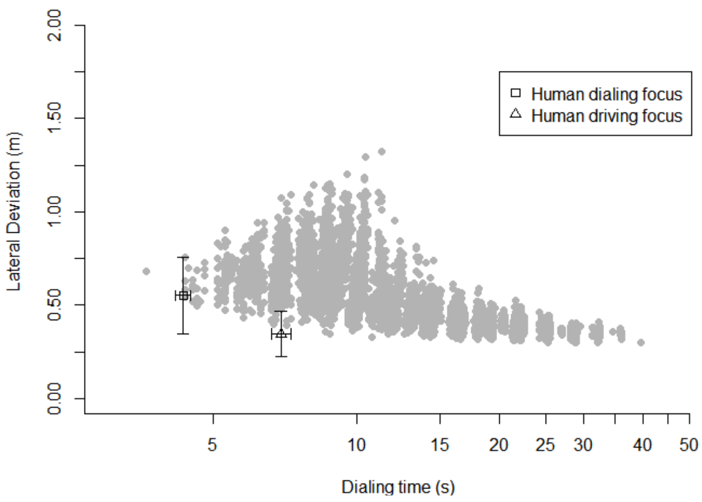

Congruent number (07333-888111) - PP 9
Lateral Deviation (m)
0.00
0.50
1.00
1.50
2.00
Human dialing focus
Human driving focus
5
10
15
20
25
30
40
50
Dialing time (s)


Congruent number (07333-888111) - PP 10
Lateral Deviation (m)
0.00
0.75
1.50
2.25
3.00
3.75
4.50
5.25
Human dialing focus
Human driving focus
5
10
15
20
25
30
40
50
Dialing time (s)

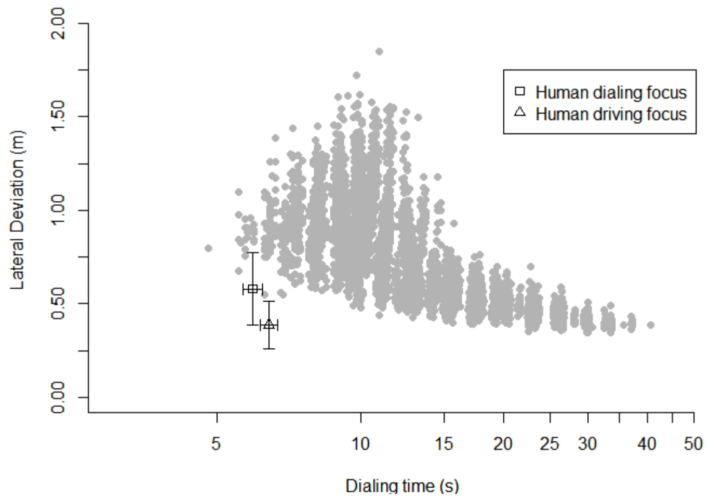

Congruent number (07333-888111) - PP 12
Lateral Deviation (m)
0.00
0.50
1.00
1.50
2.00
Human dialing focus
Human driving focus
5
10
15
20
25
30
40
50
Dialing time (s)

Appendix 2: Incongruent individual models. Error bars represent standard deviation.

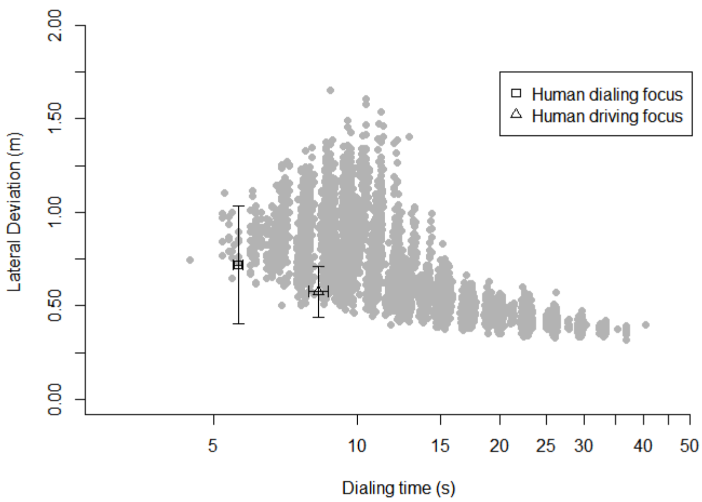
Incongruent number (07722-229944) - PP 1



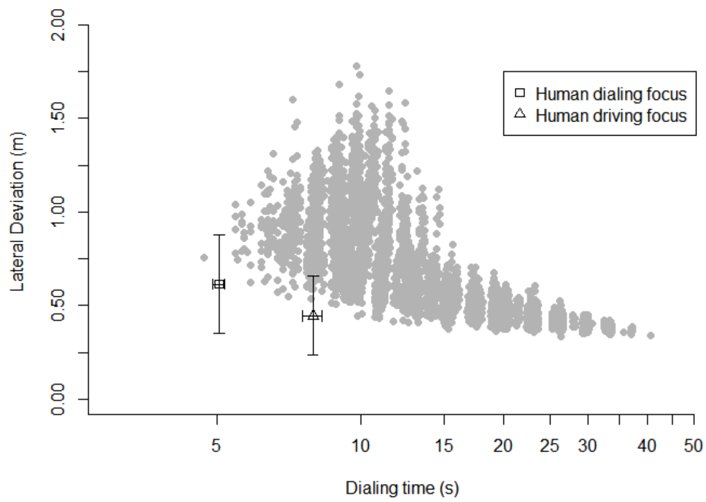
Incongruent number (07722-229944) - PP 2



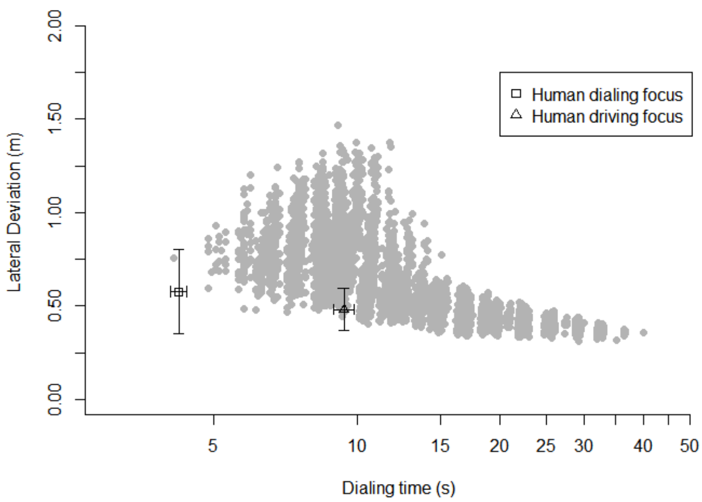
Incongruent number (07722-229944) - PP 3



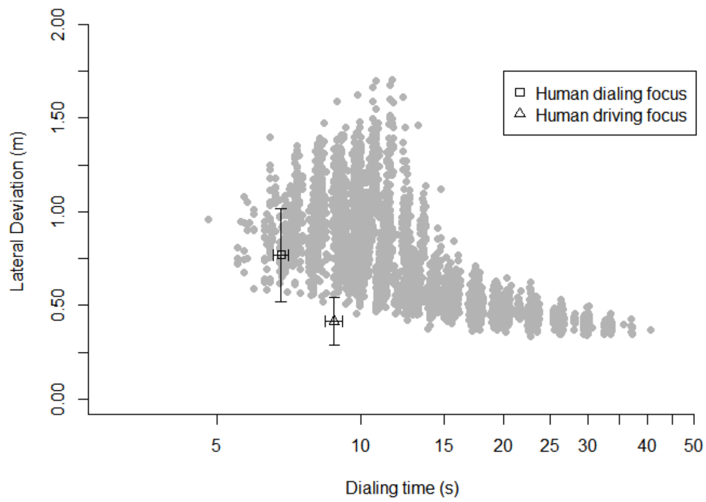
Incongruent number (07722-229944) - PP 4

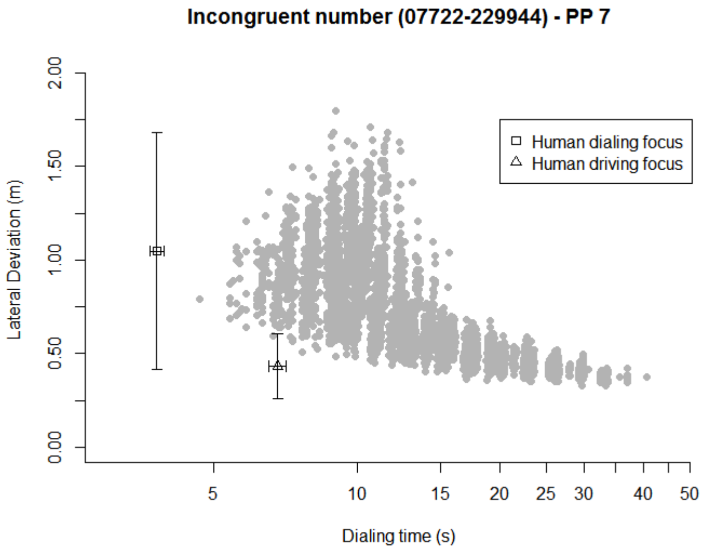


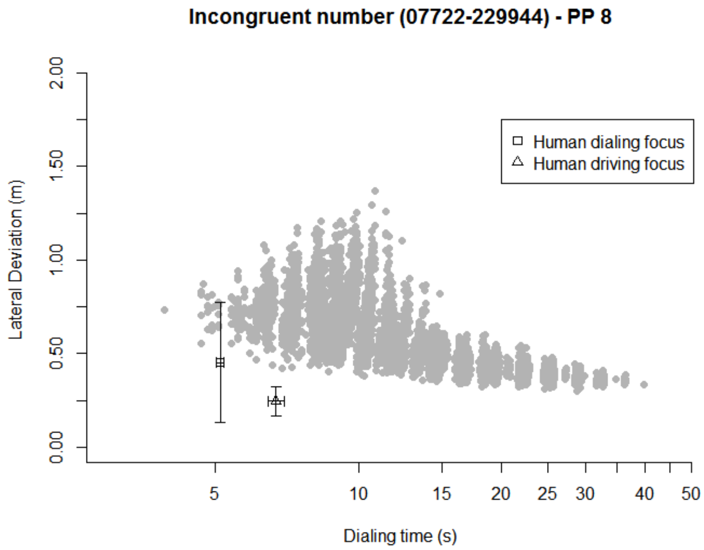
Incongruent number (07722-229944) - PP 5

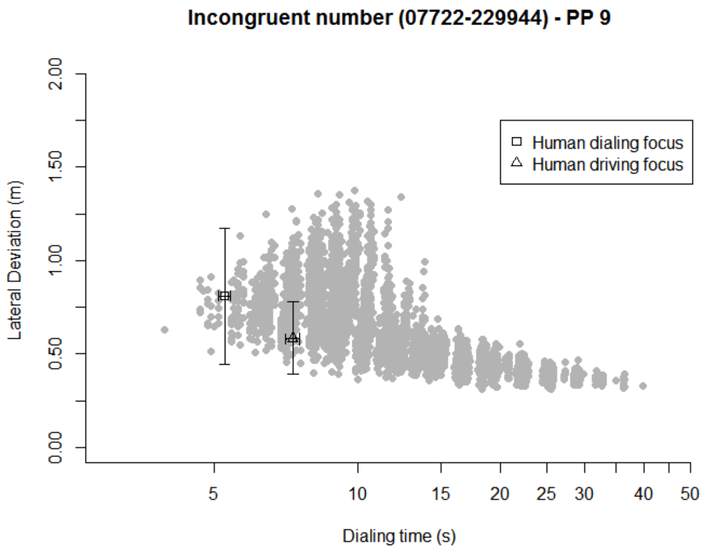


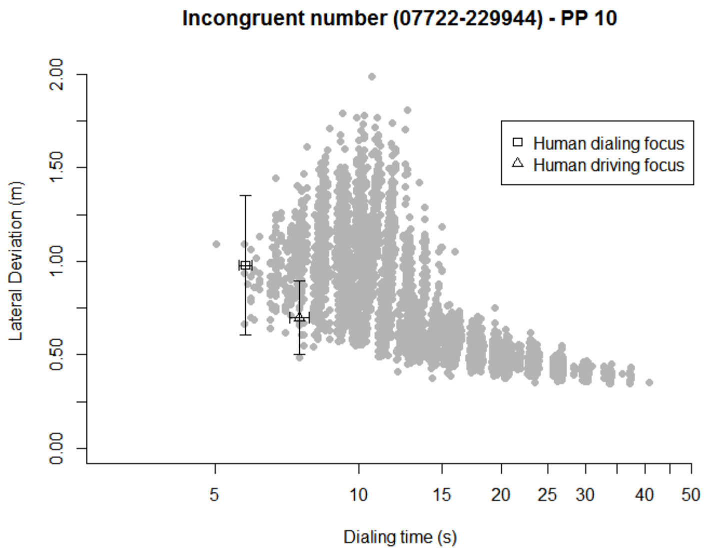
Incongruent number (07722-229944) - PP 6

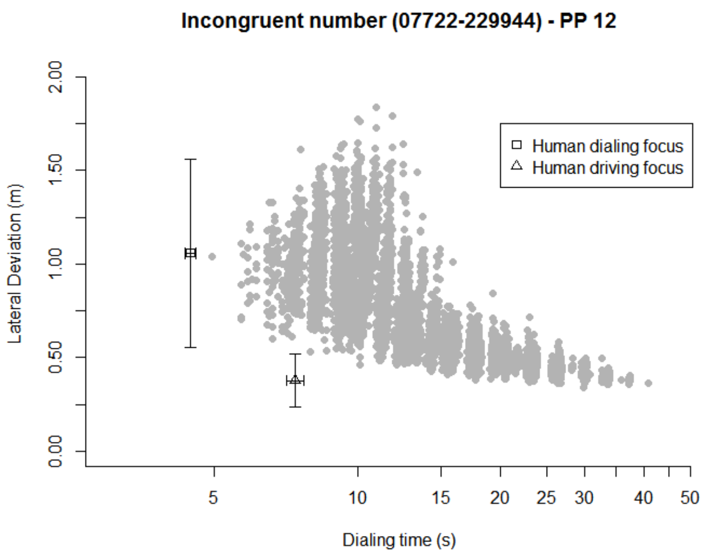



Incongruent number (07722-229944) - PP 7
Human dialing focus
Human driving focus
Lateral Deviation (m)
0.00 0.50 1.00 1.50 2.00
5 10 15 20 25 30 40 50
Dialing time (s)


Incongruent number (07722-229944) - PP 8
Human dialing focus
Human driving focus
Lateral Deviation (m)
0.00 0.50 1.00 1.50 2.00
5 10 15 20 25 30 40 50
Dialing time (s)


Incongruent number (07722-229944) - PP 9
Human dialing focus
Human driving focus
Lateral Deviation (m)
0.00 0.50 1.00 1.50 2.00
5 10 15 20 25 30 40 50
Dialing time (s)


Incongruent number (07722-229944) - PP 10
Human dialing focus
Human driving focus
Lateral Deviation (m)
0.00 0.50 1.00 1.50 2.00
5 10 15 20 25 30 40 50
Dialing time (s)


Incongruent number (07722-229944) - PP 12
Human dialing focus
Human driving focus
Lateral Deviation (m)
0.00 0.50 1.00 1.50 2.00
5 10 15 20 25 30 40 50
Dialing time (s)

Appendix 3: Congruent keypress times. Times are in ms.

07333-888111 keypress times:

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit	Second digit
1	367	390	386	688	365
2	562	501	343	360	372
3	543	353	321	555	468
4	438	626	477	824	515
5	584	441	279	351	461
6	763	640	397	451	975
7	368	497	291	573	495
8	375	604	284	397	380
9	652	662	244	598	462
10	478	387	306	3333	566
12	253	352	300	339	453

Appendix 4: Incongruent keypress times. Times are in ms.

0722-229944 keypress times: *Notice that 0722-229944 does not have a second digit since it is a first of a repeat.*

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit
1	275	345	318	360
2	878	451	362	291
3	481	513	285	489
4	452	514	340	489
5	350	484	253	566
6	424	485	375	536
7	457	576	276	517
8	299	279	404	467
9	469	485	190	556
10	379	505	374	741
12	928	414	352	592

Appendix 5: Dualtask keypress times. Times are in ms.

07333-888111 dual task keypress times, dialing focus:

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit	Second digit
1	490	553	364	137	534
2	777	686	377	395	728
3	783	667	434	267	549
4	480	404	316	225	500
5	661	606	267	265	416
6	714	723	384	750	599
7	468	474	282	327	429
8	655	702	370	256	509
9	864	868	286	364	723
10	849	845	409	424	527
12	589	580	290	260	479

07333-888111 dual task keypress times, driving focus:

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit	Second digit
1	1450	1230	383	1050	731
2	1010	1210	377	592	808
3	1420	1200	346	556	1070
4	877	646	414	540	714
5	1760	1950	450	2000	1820
6	967	1250	435	1410	1200
7	841	601	340	500	720
8	782	987	381	381	524
9	1770	1760	492	1110	1470
10	1280	981	384	660	682
12	816	893	376	348	909

0722-229944 dual task keypress times, dialing focus:

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit
1	348	559	312	171
2	360	846	363	368
3	554	639	422	396
4	380	571	351	324
5	188	571	302	225
6	705	821	377	930
7	405	482	240	254
8	561	545	387	436
9	547	654	352	324
10	0.5	727	382	461
12	478	516	304	384

0722-229944 dual task keypress times, driving focus:

Participant Nr	Chunk boundary	First of repetition	Repeating digit	Very first digit
1	378	979	409	601
2	384	894	396	480
3	1380	1050	378	815
4	743	921	391	1550
5	596	1340	433	1270
6	1110	1130	422	1070
7	663	941	372	520
8	735	875	389	521
9	783	941	426	627
10	510	1120	386	602
12	1030	946	396	592

Appendix 6: Sum of all keypress times. Times are in ms.

Congruent dialing focus:

Participant Nr	Dual-task sum	Pretest sum	Δ
1	2078	2196	-118
2	2963	2138	825
3	2700	2240	460
4	1925	2880	-955
5	2215	2116	99
6	3170	3226	-56
7	1980	2224	-244
8	2492	2040	452
9	3105	2618	487
10	3054	5070	-2016
12	2198	1697	501

Congruent driving focus:

Participant Nr	Dual-task sum	Pretest sum	Δ
1	4844	2196	2648
2	3997	2138	1859
3	4592	2240	2352
4	3191	2880	311
5	7980	2116	5864
6	5262	3226	2036
7	3002	2224	778
8	3055	2040	1015
9	6602	2618	3984
10	3987	5070	-1083
12	3342	1697	1645

Incongruent dialing focus:

Participant Nr	Dual-task sum	Pretest sum	Δ
1	1390	1298	92
2	1937	1982	-45
3	2011	1768	243
4	1626	1795	-169
5	1286	1653	-367
6	2833	1820	1013
7	1381	1826	-445
8	1929	1449	480
9	1877	1700	177
10	1570	1999	-429
12	1682	2286	-604

Incongruent driving focus:

Participant Nr	Dual-task sum	Pretest sum	Δ
1	2367	1298	1069
2	2154	1982	172
3	3623	1768	1855
4	3605	1795	1810
5	3639	1653	1986
6	3732	1820	1912
7	2496	1826	670
8	2520	1449	1071
9	2777	1700	1077
10	2618	1999	619
12	2964	2286	678

Appendix 7: Most nearby points on horizontal axis.

Abs Δ denotes the difference between the closest point of the model, and the human mean point.

Congruent average model

	Dialing focus		*Driving focus*	
Participant Nr	**Dialing Time (ms)**	**Abs Δ**	**Dialing Time (ms)**	**Abs Δ**
1	4363	88	8033	48
2	5533	4	7123	33
3	5533	6	7533	17
4	4363	454	5933	26
5	4363	207	12203	14
6	5763	48	8703	4
7	4363	499	5283	20
8	5013	34	5933	14
9	5363	42	10873	48
10	5933	12	6863	31
12	4363	137	6033	84

Incongruent average model:

	Dialing focus		*Driving focus*	
Participant Nr	**Dialing Time (ms)**	**Abs Δ**	**Dialing Time (ms)**	**Abs Δ**
1	4435	120	6935	3
2	5935	7	6435	15
3	5605	15	8255	28
4	5185	143	7935	2
5	4435	225	9345	29
6	6775	27	8775	22
7	4435	646	6775	35
8	5185	74	6685	15
9	5255	8	7195	106
10	5935	156	7505	8
12	4435	43	7435	48

Congruent individualized model:

	Dialing focus		*Driving focus*	
Participant Nr	**Dialing Time (ms)**	**Abs Δ**	**Dialing Time (ms)**	**Abs Δ**
1	4516	65	8016	31
2	5524	13	7114	24
3	5598	59	7538	22
4	5891	1982	5891	16
5	3952	204	12192	3
6	5851	40	8751	52
7	4276	412	5276	27
8	5064	17	5964	17
9	5400	5	10840	15
10	6987	1042	6987	93
12	4236	10	6086	31

Incongruent individualized model:

	Dialing focus		*Driving focus*	
Participant Nr	**Dialing Time (ms)**	**Abs Δ**	**Dialing Time (ms)**	**Abs Δ**
1	4355	40	6945	7
2	5953	25	6453	33
3	5617	3	8267	16
4	4697	345	7937	4
5	4117	93	9367	7
6	6775	27	8775	22
7	4658	869	6828	18
8	5072	39	6662	8
9	5402	155	7312	11
10	5760	19	7510	3
12	4936	458	7356	31