A Large-Scale Analysis of Attentional Deployment across One Hundred Sessions of Adaptive Multitask Training

Omar Claflin, MD PhD (omarclaflin@gmail.com)

Abstract

Human cognition is routinely challenged by today's multitasking demands which require continuous attentional deployment to multiple task components in parallel. While practice-based multitasking training has been shown to improve multitasking performance, little is known about how attention should be best deployed for optimal training. To this end, we leveraged a large-scale dataset from an online cognitive-training platform to investigate individual differences in task learning across long-term training. We developed an index of attentional deployment that specifies the temporal dynamics of learning for each component of the multitask and calculate distance maps between clusters of users to specify distinct learning styles. While long-term practice improved the multitasking performance of all participant groups, participants who focused on learning one task component earlier and more emphatically, benefited from superior learning gains throughout the entirety of training.

Keywords: multitasking; attentional bias; cognitive control; learning; adaptive training; practice effects; big data

Introduction

Human cognition is marked by the ability to task switch which allows a person to rapidly adapt to different situations. Humans regularly challenge these control processes when attempting to simultaneously integrate multiple sources of information. In tasks such as driving, human cognition adapts to input discrete information cues such as traffic signals and GPS systems along with continuous environment scanning for cars to avoid collisions or to stay within the lanes. Many jobs in the modern era include integration of information sources from discrete warning systems while simultaneously demanding continuous monitoring of other sources.

Various theoretical models of multitasking attempt to describe improvements in multitasking ability following training (Kiessel, et al, 2010). These theories range from changes in specific task representation to improvements in a more generalized cognitive flexibility ability (Altman & Gray, 2008, Dux, et al, 2009, Salvucci & Taatgen, 2008, Steyvers, 2019). According to these theories, processes that improve A) task-switching, B) individual task performance, C) a learned fused representation of both tasks, and D) learning a more generalized multitask representation, are all candidates for contributing to performance gains after training. While discussion of these sources seems purely theoretical, they have specific implications for training program structures. For instance, if gains are primarily derived from cognitive flexibility and plasticity, this would suggest that generalized representations of multitasks are the source of gains and would lead to more recommendations to immerse trainees in varied types of multitasking. While this may be the case, there is still much to be understood about the most effective learning of an individual multitask, which would form the elements in a much more comprehensive multitask training program.

For purposes of understanding the processes behind multitasking and generating usable training recommendations, we will focus on theoretical sources contributing to the performance gains within a single multitask following practice-based training. This still leaves many theoretical sources available as explanations for the observed gains in cognitive performance. For instance, if performance gains are primarily derived from strengthening single-task representations contributing to multitask performance, these theoretical sources would suggest that learning individual task components, still embedded in the multitask context, should be focused on individually. However, other models of cognitive flexibility would stress integrated cognitive representation of both tasks, which would suggest the opposite training recommendation. In this case, trainees might be best served by immediate immersion into the multitask context with feedback and direction to emphasize equal attention to learn most effectively the integrated multitask.

To investigate these cognitive accounts of multitasking, and their relevant training recommendations, we can leverage large-scale cognitive training data available from Lumosity's Human Cognition Project, an online training platform. While task instruction and feedback within these training games reinforce a balanced task emphasis across this dual-task paradigm, there likely exists significant individual variation. Of interest to researchers, the variation of task emphasis may in fact contribute to individual differences in learning effectiveness and speed. More evenly split task attention would strongly support computational models of fused tasks, while attentional bias to one task supports the notion of single-task representation models as well as single-task learning within the context of a multitask perceptive environment before task integration.

Separating the dynamics of attentional bias towards the task components provides a method to quantify the session-to-session variation that occurs during learning and a way to calculate to what extent each type of multitask learning occurs. Doing this across a large number of participants (e.g. tens of thousands) makes it possible to compare clusters of learning styles while, importantly, significantly improving the signal to noise ratio (approximately, by an order of hundred in this paper) that is normally debilitating in typical participant sized pools when examining second-order behavioral measures within participants.

Large-scale behavioral data affords us the opportunity to calculate second-order behavioral metrics of reaction time. Often, the distribution is bounded to the left and has a long right tail, popularly modelled by an ex-Gaussian distribution (Whelan, 2008). Although the model parameters and its exact mental correlates are debatable (Matzke and Wagenmakers, 2009), there is ample suggestion that at minimum, right-tailed skews indicate lapses in cognitive control or attention to the task, while median shifts indicate intrinsic speed of the processes related to the task (Dawson, 1988). We push this concept further by separating out individual task components within the adaptive multitask in order to get sensitive time dynamics of attentional emphasis for both task components across 100 learning sessions.

Separating the task components of a dual-task across longitudinal training data specifically allows us to make comparisons of ultimate cognitive performance gains, learning speeds, and the degree to which learning patterns (namely, attentional emphasis on each task component throughout learning) affects performance gains and learning speeds.

Materials and Methods

Paradigm and Participants

Adaptive Multitasking Paradigm We performed a retrospective analysis of gameplay data for a sample of participants who played the Web version of "Highway Hazards" on the Lumosity platform that is designed to test the ability to split attention between avoiding different kind of obstacles on the road (1) *continuous non-cued events*, and (2) *discrete cued events*. Trials from non-cued events (Task 1, for purposes of this paper) originate on the horizon (*Figure 1*) and successful avoidance is dependent on the current speed of the car. Trials for the secondary task are cued with statically presented signs and roll onto the road from the side beyond the visual field of view with no other warning. Time between the cue and these events (Task 2) are based on the current participant level.

Sessions have a fixed duration of 180 seconds and involve both kinds of trials. Failure to avoid either kind of these obstacles result in collisions and decrease the speed of the car. Progressively avoiding obstacles successfully results in a gradual increase in car speed. The goal of the player is to keep the road speed as high as possible (capped at a maximum speed based on the current level of the participant). At the end of each gameplay session, participants are provided feedback and the participant's level is changed for the next session depending on performance.

Sample Size Participants who played at least one hundred sessions of Highway Hazards were included in this retrospective study. Participants with greater than 1% missing data, contiguous missing data, played more than four years ago, starting at levels other than the default starting level, that contained sessions that dropped below the starting level were excluded were dropped. Linear interpolation of missing data resulted in interpolation of 0.79% of the session data. This resulted in 42,932 participants who showed learning progress over 100 sessions on this version of the game.

Data Processing

Trial Analysis Reaction time to events were calculated by subtracting the onset of the cue from the time of the first action that successfully caused avoidance of the obstacle. For the primary task (where there is no warning), reaction time was calculated as the difference between the onset of the obstacle on the horizon from the time of the action that resulted in a successful collision avoidance. Due to the adaptive nature of this paradigm, difficulty level changes were used as the performance measure and only correct trials were included for reaction time analysis.

Session Analysis Difficulty level progression was calculated by performing a within-participant normalization so that the maximum level achieved in 100 sessions for every participant was 1 and the starting level for all participants was 0. While this has no effect on the analytical method (because it relies on time-series correlations between participant progression which are unaffected by normalization), normalization allows the data view to be undominated by final level achieved thus compensating for differences in participant's ability across training in the data views (Figure 2, 3).

An attentional allocation metric was aggregated by calculating the skew of the reaction time distribution within each task type (Task 1 = *continuous non-cued*, Task 2 = *discrete cued events*) within each session. The tail of reaction time is a known correlate of attentional allocation to the task with right-tailed or positive skews indicating distraction or lapses in attention and negative tails indicating possible hypervigilance or impulsivity. Since sessions are comprised of many trials, distributions of reaction time were aggregated for each session of play for each user. The tail of the reaction time distribution for each session was estimated using a skew formula: (*mean – median*)/*standard deviation* of all reaction times within each session.

An index of Attentional Bias was calculated by (1-*Task 1 attentional allocation)/(1-Task 2 attentional allocation)* to indicate the relative bias of attentional emphasis between the two tasks during this multitasking paradigm across each session of training. Reaction time skews for each task were normalized within-participant prior to this computation to remove intrinsic task component imbalances within the game and intrinsic biases within participants. Following this procedure, we can visualize session-to-session changes within a participant. This metric is designed to reflect the relative emphasis of attention between the two tasks, by representing the extent of attentional bias away from the primary task with respect to the secondary task. Larger numbers indicate less distraction within Task 1 reaction times (resulting in smaller positive tails) in comparison to the extent of distraction within Task 2, and thus, reflects an increased attention bias towards Task 1.

Participant Analysis Difficulty level means and distributions across all participants were calculated and presented (*Figure 2)* reflecting mean ability progression across 100 sessions of training.

Attentional bias means and distributions across all participants were calculated and presented (*Figure 4*) reflecting mean task emphasis across 100 sessions of training. Peak attentional biases and session number across the 100 sessions of training were calculated to reflect the session which the average participant has a maximum bias towards the secondary task and away from the primary task (*Figure 4*).

Cluster Analysis Time-series clustering of learning behaviors was achieved using multidimensional scaling (MDS) on a distance matrix generated by using temporal representational dissimilarity matrices (RDMs) (Kriegeskorte, Mur, & Bandettini, 2008) on the first sixteen sessions of the Attentional Bias metric across all participants. A temporal region of interest (ROI) was *a priori* determined by calculating the 50% gain in the overall difficulty level gains (determined to be 16.35 sessions). Representational dissimilarities across the first sixteen sessions between all users were computed and then Fischer transformed. Temporal RDMs resulted in a 42,932 x 42,932 distance matrix between all participants. Hierarchical agglomerative clustering was used on this distance matrix, and an arbitrary distance cutoff of *0.5* of the maximum distance was used which separated the participants into fifteen learning trajectories (*Figure 5*).

Training gains, calculated as the area under the curve (AUC) of the difficulty level curve, was computed for each learning cluster (*Figure 6*) and 1-way ANOVAs were computed at sessions 20, 50, and 100. Additionally, the mean Attentional Biases across sessions were computed and ANOVA results along this clustering variable are presented. Bootstrapped distributions (without replacement) for all ANOVAs computed in this study, obtained by random assignment of participants in clusters of identical size to the analysis and were resampled 1,000 times. These bootstrapped distributions were used to compute a significance ($p \leq 0.01$) for all analyses.

Temporal Dynamic Modelling Time of peak attentional bias as a predictive measure for training gains was confirmed by a linear regression model. The session number of the peak attentional bias was used to predict the AUC of difficulty across all sessions. This target measure reflects users who achieved their maximum potential earlier and maintained it for longer. For comprehensiveness, the correlation between the AUC of the attentional bias metric and AUC of difficulty level are also reported for significance comparison.

SNR Analysis Signal-to-noise ratio across all participants was estimated by computing the absolute mean error from a linear model of Attentional Bias, using the last 80 sessions for all users (well after the peak for all clusters) resulting in an SNR of *39.3*. Using the SNR formula $SNR = E[S^2]/E[N^2]$, SNR for an individual participant can be thus estimated at *0.19*, reflecting a stronger noise than signal at a single participant level. For a SNR of at least 5, approximately 700 participants are needed. For our study, the smallest cluster had 1,881 participants.

Results

Difficulty Level Mean difficulty levels and changes in difficulty levels across all participants for all sessions were computed revealing a characteristic logistic pattern of ability gain (*Figure 2, 3*).

Attentional Bias An attentional bias index across all participants for all sessions was computed revealing an early positive peak followed by a return to mean, potentially characteristic of an early learning phase (*Figure 4*).

Clustering analysis Clustering of 43,932 participants by temporal patterns in the attentional bias index in the first 16 sessions of training using a distance cutoff of 0.5 resulted in fifteen clusters (*Figure 5*) Cluster sizes ranged from 1,881, to 5,394 participants. Mean positive peaks within the attentional bias time course for each cluster were found at sessions 0 to session 14 (*Figure 5*). Mean time of attentional bias peaks for each cluster were input into an ANOVA model confirming the clustering procedure ($F=421.3, p < 0.001$).

These clusters were then used to evaluate difficulty level progression at session 20, at a time point well after the clustering procedure but still within early training, which revealed mean difficulty level differences ($F=10.01, p < 10^{-23}$; *Figure 6*) where mean difficulty values reflect the proportion of the maximum level eventually achieved by the participant. ANOVAs for gains at session 50 ($F=5.16, p < 10^{-10}$) and session 100 ($F=2.6, p < 10^{-3}$) were also computed, reflecting diminishing differences in the learning groups as they get further away from the early learning phase, likely due to other sources of learning variance over the course of training.

Temporal Modelling Target variables of interest (time of peak attentional bias) across all participants was used to predict training effectiveness (total area under the difficulty level curve) across all 100 sessions *(slope = -0.000898, intercept = 0.62, r = -0.23)*. For comparison, AUC for both metrics were linearly regressed across all sessions for each participant (*slope = -0.0856, intercept = 0.668, r = -0.0746).*

Mean time for peak attentional biases for each of the fifteen learning clusters was extracted and used to predict training effectiveness by session 20 (*slope = -0.00185, intercept = 0.465, r = -0.82, p < .0005),* session 50 (*slope = -0.0014, intercept = 0.646, r = -0.881,* $p < 10^{-4}$*),* and session 100 (*slope = -0.001, intercept = 0.79, r = -0.73, p < 0.005)* (*Figure* 6).

General Discussion

We present behavioral evidence of individualized multitask adaptation within practice-based learning against an adaptive multitask paradigm. This study offers a group-level insight into the temporal dynamics of attention across training; these dynamics ultimately contribute to increases in multitasking ability. In addition to the aggregate training time contributing to multitask ability improvement, a pattern of attentional bias to an isolated task component early in training results in superior training speed and gains. This temporal pattern may reflect focused but effective learning of one task component, temporarily treating the other task components as noise, to effectively learn the rules of one task component well. Note that rather than true single-task practice, this single-task emphasis still takes place in multitask context where other active task components are ongoing. This distinction is important because adaptation to the individual task components also potentially involve effectively filtering out visual and motoric aspects of the other components that are present in a multitask.

Specifically, we visualized dynamics within our learning clusters of contiguous sessions of focus on a task component, marked by an above baseline attentional focus, anywhere from one to six sessions, early in training, which seems to constitute a necessary learning phase to adapt to the multitask. However, those participants with less biased attentional deployment towards one task component, or who deployed biased attentional deployment later in training, ended up with the least gains and the slowest training speed. Counter to intuition derived from many multitasking models which might stress multitask integration and a more balanced attentional load to all components of the task, users that frontloaded the most attentional bias towards one task component within the earliest session or sessions, benefitted from more superior training gains and multitask performance throughout training.

It is important to note that rather than reflecting a learning style to focus on the individual components, this pattern may reflect cohorts that are able to learn faster, reflecting those that could achieve higher peaks of attentional bias and spend less sessions on that learning phase. As this data shows, those faster learners could thusly achieve their maximum potential ability earlier. In either case, we demonstrate this as a useful metric of an early learning phase of multi-session multi-task adaptation.

Our second-order attentional metric and time-series clustering procedure generates simple canonical functions of attentional dynamics across the earliest sessions of training that represent different learning styles which ultimately contribute to differences in individual user adaptation to an adaptive multitask. This approach had the advantage of not requiring *a priori* models, while also being a clustering approach that provides descriptive predictive features (in the form of the canonical functions).

By generating temporal canonical functions which emerge from large datasets, this paper provides a possible direction in the future by which individual or group variation within mass-collected behavioral datasets can be used to study cognitive adaptation over time. Future beneficial directions include improving reaction time with more sophisticated ex-gaussian models or other distributions, using smoothing filters to improve RDM measures, increasing the dimensions of the MDS map, using hyperparameter searches to find optimal temporal ROIs and optimal number of clusters. Different clustering thresholds ranging from five clusters to thirty clusters resulted in similar findings but were not reported due to the limitations of this format. Future studies may wish to include using the described methods in this paper for analyzing temporal dynamics of action efficiency, reactivity versus planning biases, recovery from errors, and other reaction time based components of multitasking, in the context of practice-based adaptation.

Finally, future multitasking training programs may be made potentially more effective by providing direction to trainees to isolate and focus attentional emphasis upon a single task component, in the context of a multitask environment, early in training, before proceeding to improve general multitask integration.

Figures

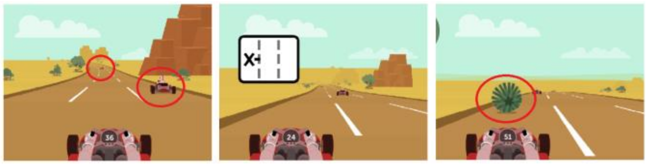

Figure 1: Web-based version of *Highway Hazards. Left:* Representative obstacles from Task 1 where the goal is primarily to avoid collisions with cars which appear on the horizon and get closer (red circles digitally added for clarity and are not present in the paradigm). *Middle:* Cue event for Task 2 which precedes an obstacle. *Right:* Obstacle for Task 2 that appears from the side of the road.

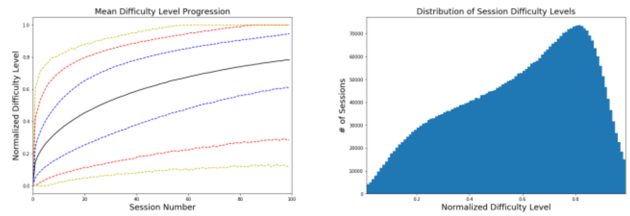

Figure 2: Difficulty levels across all sessions and participants. *Left:* Mean within-participant-normalized difficulty level for all users across 100 sessions of training (mean is *black*; *blue, red, yellow* lines reflect the 68%, 95%, and 99% confidence intervals). *Right:* Distribution of within-participant-normalized difficulty level for all sessions.

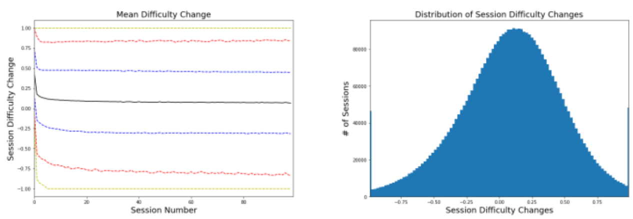

Figure 3: Difficulty changes (normalized within participants) across all sessions and participants *Left:* Mean training changes (in *black*) for all users across 100 sessions of training (*blue, red, yellow* lines reflect the 68%, 95%, and 99% confidence intervals). *Right:* Distribution of difficulty changes across all training sessions.

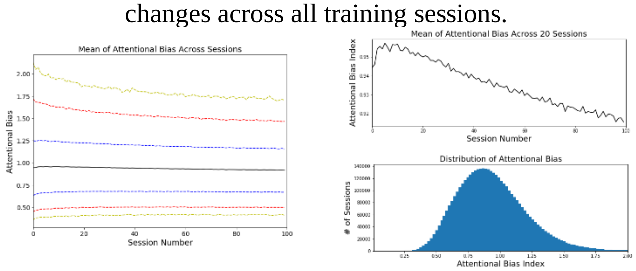

Figure 4: Attentional bias across all sessions and participants. Greater numbers reflect increased attention towards the primary task, relative to the secondary task. *Left*: Mean Attentional Bias (in *black*) for all users across 100 sessions of training (*blue, red, yellow* lines reflect the 68%, 95%, and 99% confidence intervals). *Right top:* For clarity, mean is plotted without confidence intervals, to reflect the positive peaks of attentional deployment that occur at session 5 and session 10. This is followed by a steady reduction in attentional deployment towards Task 1 over 100 sessions of training. *Right bottom*: Distribution of Attentional Bias across all training sessions.

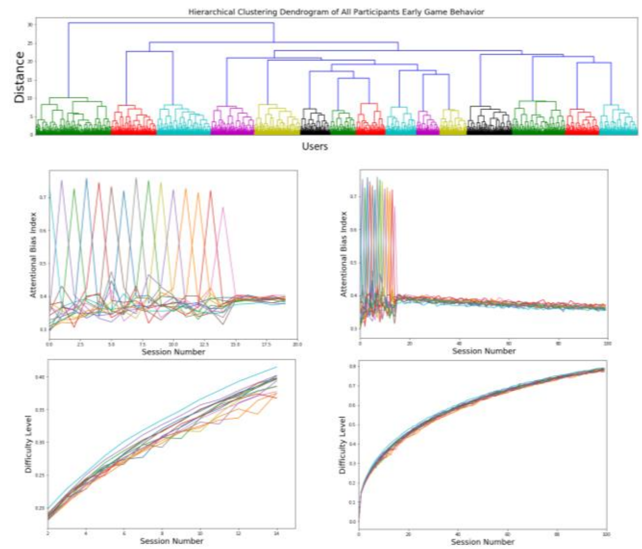

Figure 5: Time-series RDM-based clustering of attentional bias in early training sessions. *Top:* Clustering of 43,932 participants by temporal patterns in attentional bias behavior in the first 16 sessions of training. *Middle left*: Mean attentional bias index across these clusters for the first 16 sessions. Note the positive peaks in attentional deployment occurring in each cluster (clarified in Figure 6). *Middle right:* Mean attentional bias time course across these clusters for all 100 sessions. *Bottom left*: Mean difficulty level progression within the first 16 sessions for these clusters. *Bottom right*: Mean difficulty level progression across all 100 sessions for these clusters.

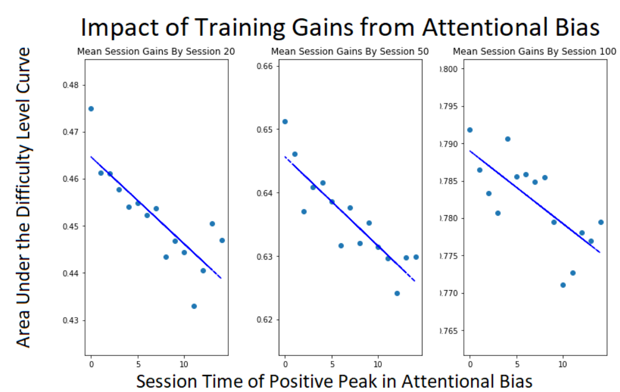

Figure 6: Attentional bias and total progression (measured as area under the curve) of the fifteen different participant clusters. Earlier peaks in attentional bias strongly correlated with increased training gains across the entirety of training ($p < 0.005$). Training gains at Session 20, 50, and 100 (*left*, *middle*, *right*) are presented.

Acknowledgments

I would like to thank AF, AO, NN, RS, AK, and OJ at Lumos Labs for their input, BKO, and the anonymous reviewers for their thoughtful comments. Thank you to The Human Cognition Project (https://www.lumosity.com/hcp) for the data.

References

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological Review, 115*(3), 602–639.

Dawson, M. R. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments & Computers, 20*(1), 54–57.

Dux, P. E., Tombu, M. N., Harrison, S., Rogers, B. P., Tong, F., & Marois, R. (2009). Training improves multitasking performance by increasing the speed of information processing in human prefrontal cortex. *Neuron, 63*(1), 127–138.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136*(5), 849–874.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4.

Lussier, M., Gagnon, C., & Bherer, L. (2012). An investigation of response and stimulus modality transfer effects after dual-task training in younger and older. *Frontiers in human neuroscience*, *6*, 129.

Matzke, D., and Wagenmakers, E.-J. (2009). Psychological interpretation of the ex–Gaussian and shifted Wald parameters: a diffusion model analysis. *Psychon. Bull. Rev*. 16, 798–817.

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

Steyvers, M., Hawkins, G. E., Karayanidis, F., & Brown, S. D. (2019). A large-scale analysis of task switching practice effects across the lifespan. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *116*(36), 17735–17740.

Whelan, Robert. "Effective analysis of reaction time data." *The Psychological Record*, vol. 58, no. 3, 2008, p. 475+.