Assessing Military Perceptual Expertise with Drift Diffusion Modeling

Webb Stacy, Jeff Beaubien, Sterling Wiggins
Aptima, Inc.
Woburn, MA and Fairborn, OH
wstacy@aptima.com, jbeaubien@aptima.com, swiggins@aptima.com

ABSTRACT

One way to assess military perceptual expertise is to present brief stimuli to military experts and novices and to ask them to make a 2-choice expertise-related judgment as quickly and accurately as they can. The differences in the distributions of their reaction times and errors can then reveal differences in their perceptual expertise. The difficulty is that reaction times and errors are often not independent. In some cases, participants make a speed-for-accuracy tradeoff. In other cases, correct responses occur quickly but error responses require a lot of perceptual processing. Either way, an analysis of reaction times or errors by themselves can produce misleading results.

An increasingly popular solution is to use reaction time and error distributions together to perform a Drift Diffusion Model (DDM) analysis. This approach yields a profile of several cognitively meaningful components, including an estimate of the speed of processing, the level of the response threshold (indicating how much information the participants needed in order to make a perceptual decision), and the amount non-decision time, which often translates to the amount of time it takes to encode the stimulus. By comparison, traditional methods for assessing reaction time and accuracy (separately) do not differentiate the component cognitive processes of encoding, decision-making, and response execution, respectively. These components can yield surprising results. For example, DDM has shown that the main culprits in aging are increases in response threshold and in non-decision time but not speed of processing. Moreover, speed of processing, but not response threshold or non-decision time, is related to working memory capacity and reasoning ability in adults of all ages.

In this paper we will discuss strategy and techniques of DDM analysis. We will then illustrate them by discussing an experiment we performed with military experts and novices using this methodology, with the aim of encouraging other researchers to adapt the approach for their own research issues.

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect the official position of the Department of Defense or its components or the organizations with which the authors are affiliated.

ABOUT THE AUTHORS

Dr. Webb Stacy is a Corporate Fellow at Aptima, Inc., where he is responsible for enhancing Aptima's technology portfolio. Dr. Stacy has an interest in using modern Cognitive Science to improve experiential training. His recent work includes investigating the relationship of simulator fidelity to training effectiveness and developing an approach to optimizing the training value of experiential scenarios. Dr. Stacy holds a Ph.D. in Cognitive Science from SUNY/Buffalo, and a B.A. in Psychology from the University of Michigan, and has served as Program Co-Chair for the Society for Behavior Representation in Modeling and Simulation.

Dr. Jeff M. Beaubien is a Principal Scientist at Aptima, Inc. For the past 15 years, his work has focused largely on training and assessing leadership, teamwork, and decision-making skills. His work has been sponsored by the U.S. Navy, the U.S. Army, the U.S. Air Force, the Federal Aviation Administration, and the Telemedicine and Advanced Research Technologies Center, among others. Dr. Beaubien holds a Ph.D. in Industrial and Organizational Psychology from George Mason University, an M.A. in Industrial and Organizational Psychology from the University of New Haven, and a B.A. in Psychology from the University of Rhode Island.

Sterling Wiggins is a Principal Scientist at Aptima, Inc. He leads several projects creating technology and training to

support humans working in high risk, safety-critical environments. His research interests include Live, Virtual, and Constructive training, human-automation interaction, and adaptive aiding. He holds an M.A in Education with a focus on learning, design, and technology from Stanford University.

Assessing Military Perceptual Expertise with Drift Diffusion Modeling

Webb Stacy, Jeff Beaubien, Sterling Wiggins
Aptima, Inc.
Woburn, MA and Fairborn, OH
wstacy@aptima.com, jbeaubien@aptima.com, swiggins@aptima.com

Experts make better decisions than novices under time pressure, especially if those decisions involve quickly perceiving and reacting to situational cues. In medicine, expert surgeons quickly and accurately identify tumors to be excised; in sports, expert soccer goalies quickly and accurately recognize the likely trajectory of an opponent's penalty shot, though their opponent may also be skilled enough to deceive the goalie; and in aviation, expert pilots quickly and accurately recognize and act on significant deviations from nominal landing conditions. Do these experts more quickly encode the situation than novices? Do they process the incoming stream of information faster? Do they gather more or less information before making a decision? This paper aims to explain methods to develop answers to these questions, in an attempt to characterize the nature of perceptual expertise, especially in military domains.

One way to assess military perceptual expertise is to present brief stimuli to military experts and novices and to ask them to make a 2-choice expertise-related judgment about them as quickly and accurately as they can. The differences in the distributions of their reaction times and errors can then reveal differences in their perceptual expertise. The difficulty is that reaction times and errors are often not independent. In some cases, participants make a speed-for-accuracy tradeoff. In other cases, correct responses occur quickly but error responses require a lot of perceptual processing. Either way, an analysis of reaction times or of errors by themselves can produce misleading results.

An increasingly popular solution is to use reaction time and error distributions together to perform a Drift Diffusion Modeling (DDM) analysis. This approach yields a profile of several cognitively meaningful components, including an estimate of the speed of processing, the level of the response threshold (indicating how much information the participants needed in order to make a perceptual decision), and the amount non-decision time, which often translates to the amount of time it takes to encode the stimulus.

These components can yield surprising results. For example, DDM has shown that the main culprits in aging are increases in response threshold and in non-decision time, but not speed of processing. Moreover, the speed of processing, but not response threshold or non-decision time, is related to working memory capacity and reasoning ability in adults of all ages.

In this paper we discuss strategy and techniques of DDM analysis and illustrate them by discussing an experiment we performed with military experts and novices using this methodology. Our aim is to encourage other researchers to adapt the approach for their own research issues.

THE CHALLENGE

Reaction time (RT) can be a sensitive measure of the cognitive processes involved in perception, learning, memory, decision-making, and a variety of other domains. Typically, some stimulus is presented to the participants, and they are asked to make a decision about it, and the time between the stimulus onset and the decision is measured. Since most often the participant's choices can be scored as correct or incorrect, the experiment has two intertwined dependent variables: RT and error rate. For decades, the way that this was handled was to report mean correct RT, and to hope that error rates didn't vary much by condition, which would mean that they were an unimportant and minor source of noise in the experimental outcome.

Unfortunately, this can lead to misleading results. RTs for error responses can vary drastically from RTs for correct responses, and this fact, coupled with different error rates in different experimental conditions (such as the difficulty of the experimental stimuli or the specific instructions that are provided to the participants), can lead to a confusing picture. For example, one well-known phenomenon is the speed-for-accuracy tradeoff, where participants are willing to be less accurate but faster in some conditions and to be more accurate but slower in others. The speed-for-accuracy tradeoff generally occurs when decisions are reasonably easy for the participants—it is not a universal phenomenon. When decisions are hard, error responses can actually be slower than correct responses (Luce, 1986.)

AN INCREASINGLY POPULAR SOLUTION: DRIFT DIFFUSION MODELS (DDM)

Thus, unless error rates and RTs are consistent across conditions, analysis of correct RTs by themselves, or of errors by themselves, is unlikely to help us understand the results. To solve this problem, we turn to an increasingly popular analysis approach that takes all the data—correct RTs, error RTs, and error rates—into account. It is called diffusion analysis (Ratcliff, 1985; Ratcliff & Rouder, 1998; Ratcliff & Turlinckx, 2002.) This analysis approach has been applied to data from variety of experiments including signal detection (Ratcliff, Thapar, & McKoon, 2001), visual search (Strayer & Kramer, 1994), and perceptual judgment (Eastman, Stankiewicz, & Huk, 2007).

The basic idea of the diffusion model, illustrated in Figure 1, is that the time to make a decision in a 2-choice reaction time experiment involves quickly gathering information about which response is best until an information threshold is reached, at which point the decision is made. The information-gathering process is characterized as a random walk, and in the approach described here there are three parameters of interest that characterize it. [1]
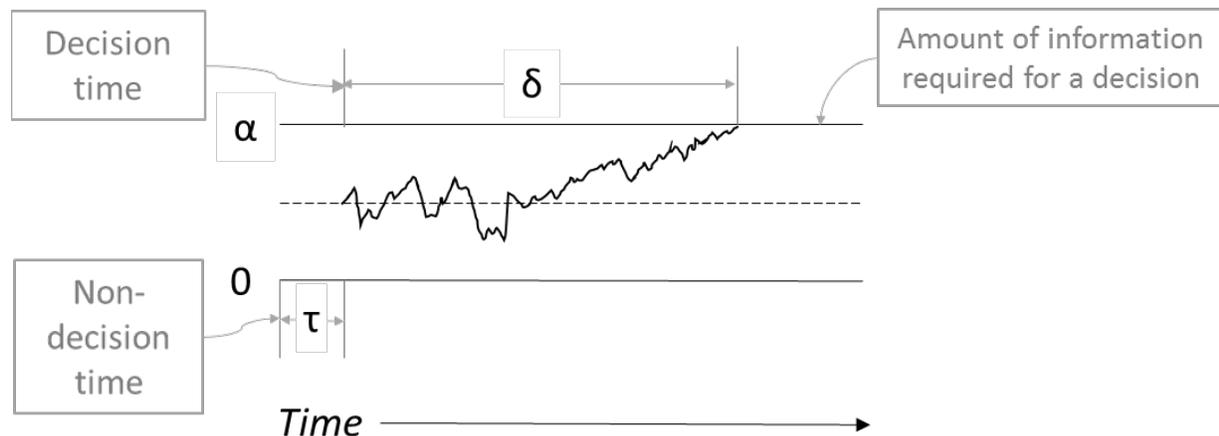

Figure 1. Elements of the Diffusion Model of Choice Reaction Time. δ, or speed of processing, is an underlying latent variable that depends on the participant's skill level and the difficulty of the task; α describes the amount of information that the participant requires in order to make a decision, sometimes called the decision criterion; and τ is the amount of time that the participant spends in activities other than decision-making, such as stimulus encoding or response execution. When response execution is trivial (such as a key press) and the same across conditions, differences in τ primarily reflect stimulus encoding time.

The three parameters, illustrated in Figure 1 1 are: 1) δ (delta): Delta measures the speed of information gathering, regardless of whether the decision was correct or incorrect. Higher values of delta mean faster decision-making, and lower values mean slower decision-making; 2) α (alpha): Alpha measures the amount of information that needs to be gathered for a decision, independent of the speed of decision-making. Higher values of alpha mean that more information must be gathered before a decision can be made; that is, higher values of alpha lead to a lower tolerance for uncertainty. Lower values of alpha mean less information is required, a higher tolerance for uncertainty, and; 3) τ (tau): Tau represent the amount of time spent on non-decision tasks like stimulus encoding and response planning and execution. Larger tau means more time is spent in these activities, and smaller tau means less time is spent. Unlike delta and alpha, which have arbitrary units, the units for tau are units of time, typically fractions of seconds. Mathematically, the diffusion model is a variant of a Wiener process (Doob, 1954), and this means that estimating its parameters is usually best done with the help of existing tools, some of which are described shortly.

These DDM parameters can be used to understand the contributions of distinct cognitive components to a wide array of cognitive phenomena. For example, a number of studies (for example, Ratcliff, Thapar, & McKoon, 2010) showed that, contrary to expectations, the mental slowdown in aging is due to changes in the amount of information required to make a decision (α) and non-decision time (τ) but not to inherent speed of processing (δ). They also showed the opposite effects with IQ, which was strongly correlated with speed of processing but not the amount of information

---

[1] There are many variants of the diffusion model, sometimes with many more parameters, but that level of complexity is beyond the scope of this paper. A fourth DDM parameter, β, or response bias, is sometimes added to the mix. It represents the participants' tendency to respond with one response more frequently than the other. It seldom provides insight into cognitive processes, and in the experiment to be described below, as in many other experiments, bias estimates hover right around .5, indicating that participants do not have a particular bias. For these reasons, β will not be discussed further in this paper.

required or non-decision time. Sleep deprivation (Ratcliff & van Dongen, 2009) and alcohol consumption (van Ravenzwaaij, Dutilh, & Wagenmakers, 2012), on the other hand, affected speed of processing but not the amount of information required to make a decision or non-decision time. In a clinical setting, White et al. (2010) showed that highly anxious individuals showed higher speed of processing (δ) for threatening words compared to nonthreatening words and compared to normal controls. DDM analyses are being performed in a variety of other domains, including Cognitive Neuroscience, as described in Forstmann, Ratcliff, & Wagenmakers' (2016) excellent review.

There are strong similarities between the DDM and Signal Detection Theory (SDT; cf. Swets, Tanner, & Birdsall, 1961; Swets, 1964; Green & Swets, 1966). Like DDM, SDT is concerned with human decision-making under uncertainty, but unlike DDM, its main focus is accuracy under different criteria for the decision rather than on both accuracy and RT. SDT calls the separation between pure noise, on the one hand, and signal + noise, on the other, sensitivity, or d'. d' is analogous to the δ parameter in DDM; both are concerned with the underlying accuracy of decisions. The other key parameter in SDT is a decision criterion, or β. β in SDT specifies the tradeoff between hits and false alarms that the decision-maker has adopted. It is most analogous to the α parameter in DDM: both parameters describe the certainty required by observers before the will make a decision. In sum: both SDT and DDM describe parameters that affect decision making; both separate underlying accuracy from decision criteria; and both concern the accuracy of decisions; but only DDM also concerns the speed with which the decision is made.

There are currently three approaches (with corresponding packages in R and other languages) that make estimation of these parameters possible. The first, EZ-diffusion (Wagenmakers, van der Maas, & Grasman, 2007) provides a simple, closed-form solution by simplifying the model. It assumes no trial-to-trial variability in speed of processing, non-decision time, and bias, and although it is not good at capturing patterns of fast and slow errors, it can produce surprisingly powerful results when simply looking for the presence of an effect (van Ravenzwaaig, Donkin, and Vanderkerckhove, 2017.) To estimate parameters for a richer version of DDM, Wabersich & Vanderkerckhove (2014a) provide Wiener distribution models and guidance about how they can be used for parameter estimation, though such use requires nonlinear optimization routines. These, in turn, can be used using traditional Null Hypothesis Significance Testing (NHST) to determine p-values and significance levels. However, as discussed below, a Bayesian analysis can provide a richer and more meaningful set of results. Fortunately, there is an equivalent package from the same authors for Just Another Gibbs Sampler (JAGS; Plummer, 2003), a tool that can approximate Bayesian posterior distributions given prior probabilities and data, that provides the Wiener distribution (Wabersich & Vanderkerckhove, 2014b).

It is important to note that, except for the EZ-diffusion approach, which is appropriate for only a subset of DDM analyses, the parameters—in this paper, we focus on δ, α, and τ—*cannot be computed by a simple closed-form equation*, the way that, say, sample means and standard deviations can be computed for data drawn from a Normal distribution. This is what generates the requirement for nonlinear optimization methods (or for Gibbs sampling, in the Bayesian case) to estimate the parameters. Further, while α and δ are defined in terms of *information units* or *rate of information units per unit time*, there is no easy way to understand the actual size of an information unit in these analyses. In contrast, τ is measured in the time units from the experiment, which in this case is seconds.

In the experiment described below, our analysis primarily used the Bayesian approach. We will not dwell on the merits of the Bayesian approach as compared to the more traditional Null Hypothesis Significance Testing (NHST) approach. We will only say that in circumstances where it is important to extract as much information as possible from the data, especially when there is prior knowledge about the answers to the questions being asked, there are advantages to the Bayesian approach. In what follows, instead of reporting NHST-style p-values, we will instead report the Bayesian posterior distribution certainty. This is a number between 0 and 1 that represents the probability mass above (or below) the "no-effect" point. Higher values represent greater certainty that there was an effect, and lower values represent less certainty that there was an effect.
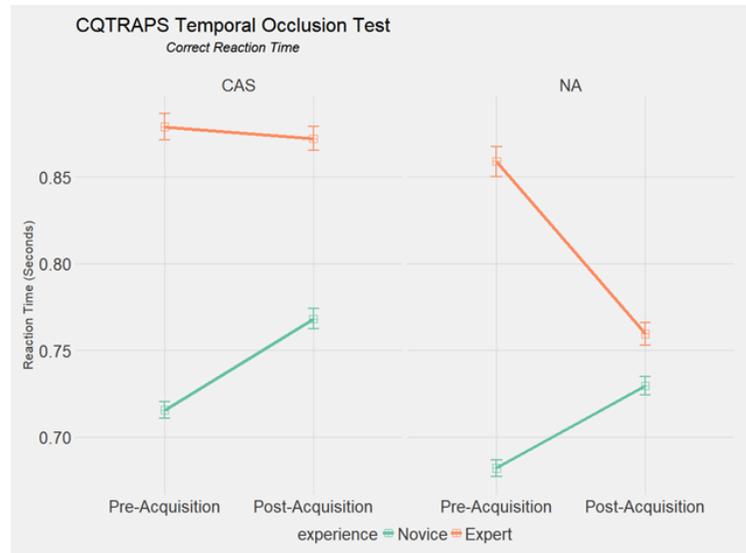
USE CASE: CARRIER LANDING EXPERTISE

Carrier landing skills have large perceptual and motor components. As it turns out, a lot of excellent work on perceptual motor expertise comes from the sports psychology community (Ward et al., 2008; Muller & Abernathy, 2012, Stacy et al., 2014). One increasingly popular set of paradigms in sports psychology include temporal and spatial occlusion (Fadde, 2010; Farrow, Abernathy, & Jackson, 2005). In temporal occlusion, participants are shown a video clip of a sports situation that will require a quick perceptual motor response, but the clip is truncated before the

situation resolves. For example, temporal occlusion training in tennis might use a video of an opponent up to the point where they strike the ball, and in baseball it might use a video of a pitcher while he winds up but before he releases the ball. Participants are then asked to "fill in the temporal blanks" and judge how the situation resolves (for instance, where the tennis ball will land, or what kind of pitch the baseball pitcher is going to throw.) In this experiment, we used a temporal occlusion technique to assess the perceptual components of carrier landing skills in US Navy F/A-18 pilots.

As part of an experiment to assess the effects of a new technology called Precision Landing Mode (PLM)[2] on carrier landing skill acquisition and decay, pilots were administered a Temporal Occlusion Test (TOT) outside the simulator, before and after in-simulator training with either conventional landing technology (Control Augmentation System; CAS) or PLM. In addition to carrier landing technology (CAS or PLM), the independent variables for the TOT included level of expertise (Novice or Expert) and whether the TOT was administered before or after the training (Pre-Acquisition or Post-Acquisition). Stimuli were created from the same video sources used in the actual simulator, and were presented in a random order controlled by Open Sesame (Mathôt, Schreij, & Theeuwes, 2012), an open-source PC-based experimental test harness. The same set of video-based situations were used for all conditions, with the main difference in stimuli being that the heads -up display (HUD) looks different under CAS and under PLM. Thus, the appearance of the HUD is a key factor in the perceptual patterns involved in the TOT experiment.

(a)

(b)

Figure 2. Correct RT (a) and error rates (b) for experimental conditions. There is no simple pattern that would allow either correct RT or error rates alone to describe the results.

The sample included 14 active-duty US Navy F/A-18 pilots. We categorized the seven participants who had 75 or fewer carrier landings as novices, and the seven who had had greater than 200 carrier landings as experts. All participants were male. At the time the experiment was conducted, PLM had not yet been operationally deployed. As a result, all of the participants – both novice and experts – were novices when it came to using PLM, regardless of their overall carrier landing experience using CAS.

The TOT itself was a computer-based perceptual skills test that was administered outside the simulator using a PC laptop and an external 22" LCD monitor. During each decision trial, the pilots were shown 8-second videos of a landing situation somewhere in a daytime pass, and were asked to indicate as quickly and accurately as possible if the situation required a standard or an aggressive correction. Put succinctly, this was a test of how quickly and how well

---

[2] Formerly called MAGIC CARPET.

the pilots perceived a troublesome landing situation.

In the TOT, there was no speed-for-accuracy tradeoff. A speed-for-accuracy tradeoff implies that error RTs are faster than correct RTs, but, as Figure 3 shows, error RTs tend to be slower.

This can happen when the task is difficult (Luce, 1986). Further evidence that the task is difficult can be seen by looking at the values of the error rates in Figure 2(b) which are in the range of 20-40%, where 50% would be chance performance. One potential explanation for how this happens is that pilots attempt to answer the question as best they can but set a deadline for responding; then, if they don't have a definitive answer when the deadline arrives, they take their most reasonable guess.

Diffusion Model Results for the TOT

Delta – Speed of Processing

Figure 4 shows the speed of processing results. Higher scores representing faster processing, (higher speed) and therefore represent better performance. As expected, experts had a higher speed of processing on CAS, but they also had a lower speed of processing on PLM. The Bayesian posterior distribution showed a 96.1% probability that the effect was greater than 0. This indicates that the perceptual pattern recognition of experts under CAS does not necessarily transfer to new technologies—in fact, there appeared to be a need to do some unlearning of those perceptual patterns under PLM, because the experts had a lower speed of processing under PLM, although with training they caught up to the novices.
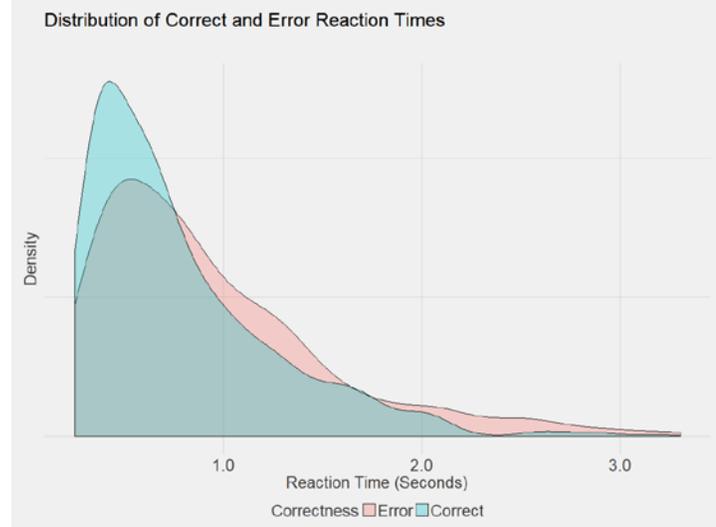

Figure 3. Distributions of correct and error RT. The graph shows kernel density estimates (basically, a smoothed histogram) for RTs of both kinds of responses. Error RTs tended to be slower and had a fatter tail, indicating there was no speed-for-accuracy tradeoff; that is, there was no speed advantage for error responses.
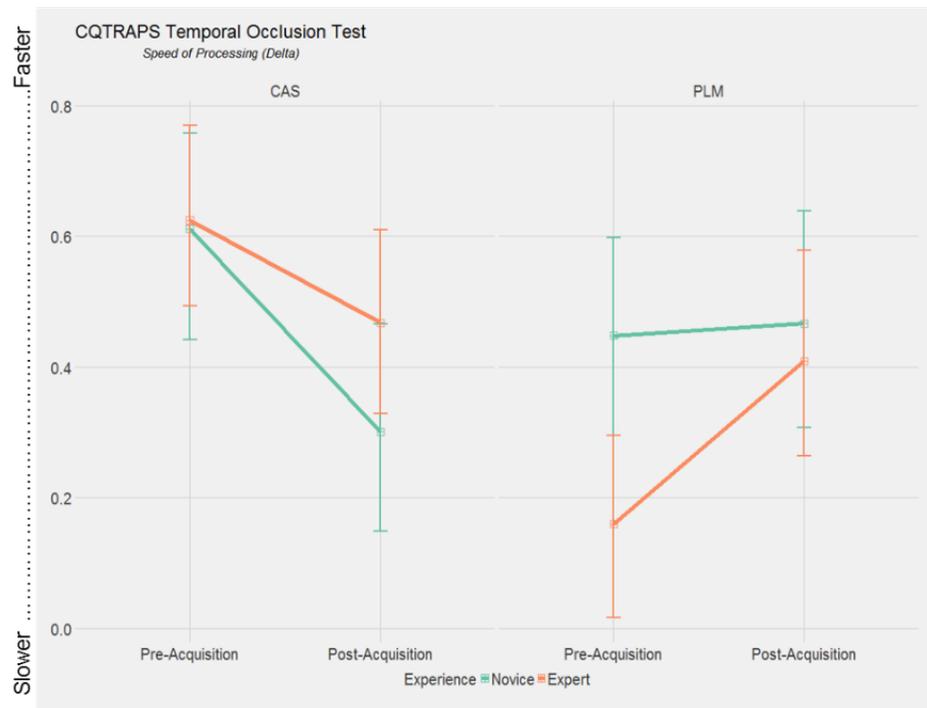

Figure 4. Speed of processing (δ) results. Speed of processing was in general higher for experts under conventional technology (CAS), but lower for experts with the new PLM technology, likely suggesting the need to unlearn recognition of some patterns. Experts caught up to novices on PLM after training.

Alpha – Amount of Information Required to Make a Decision

For α, higher scores mean that more information was required to make a decision. In the graphs, α is labelled 'Decision Criterion.' There were several interesting effects, which can be seen in Figure 5. In particular, the amount of information required to make perceptual decisions under CAS was higher than the amount required for PLM. (Bayes posterior distribution certainty 97.9%) and practice lowered the amount of information required (Bayes posterior distribution certainty 84.9%) In general, experts required more information than novices, especially for CAS (99.7% and

Figure 5. Decision criterion (α) results. In all cases, experts required more information to make a perceptual decision than novices. For CAS, experts may have been making more use of the time that their faster speed of processing enabled. It appears that the experts learned that they did not need to gather as much information for PLM after PLM training.

96.7% Bayes posterior distribution certainty for CAS and PLM, respectively), though expert information requirements shrunk under PLM practice, almost to the level of novice information requirements (90.8% Bayes posterior distribution certainty.)
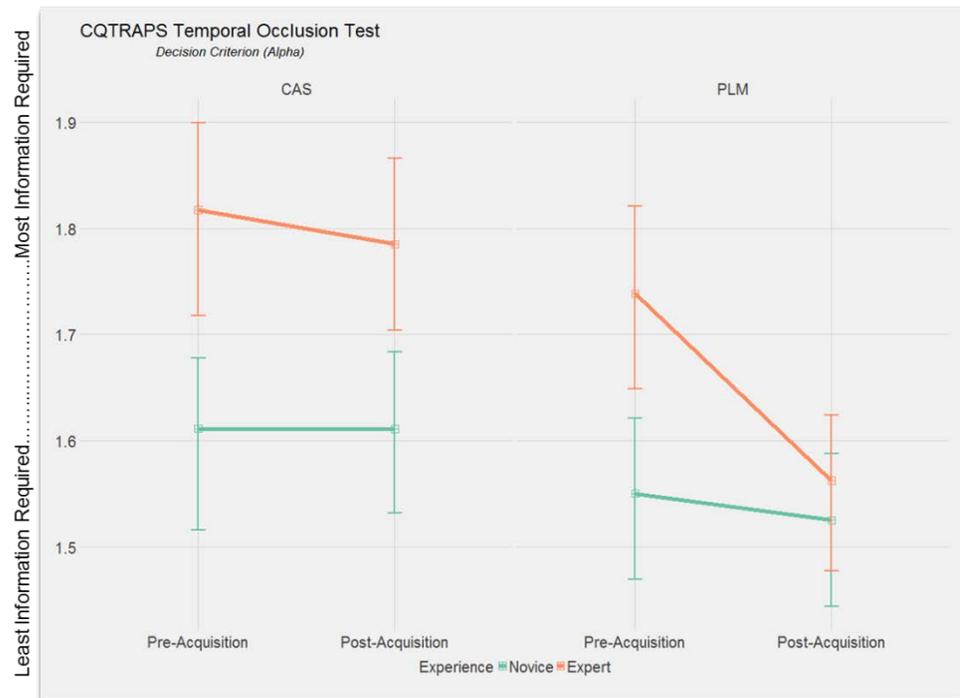
One explanation for experts needing relatively more information is that they know more about what could potentially go wrong, and thus evaluating the situations involves a more extended search than it does for novices, that is, experts are faster but more thorough when evaluating perceptual patterns under CAS.

Tau – Non-decision Time

Finally, τ is a measure of non-decision time which is measured in milliseconds. There were several interesting results in the τ analyses that indicate that experts are faster than novices at encoding stimuli from the CAS landing videos. Although tau reflects all non-decision time, not just stimulus encoding, stimulus encoding is a significant part of it (the other major part being response planning and execution, which very likely did not differ very much between participants and conditions because it only required pressing a mouse button). As Figure 6 shows, experts had less non-decision time than Novices (94.2% Bayesian posterior distribution probability), especially for CAS (96.7% Bayesian posterior distribution probability), and CAS required less non-decision time than PLM (87.4% Bayesian posterior distribution probability).

SUMMARY

The diffusion analysis suggests that with their extensive experience, experts have knowledge of CAS-specific perceptual patterns (based on the Heads Up Display) that is embedded enough in their perceptual systems that they can encode situations quickly and that their decision-making process is faster, even though they likely take more

possibilities into account. Novices, who do have some, but not as much, CAS experience, do not have these advantages, but they also have less to perceptually unlearn when switching to PLM.

Although the carrier landing performance data were not presented in this paper, PLM leads to better landing performance, even for CAS experts; and in most cases experts can quickly catch up in any case. However, expertise does not appear to fully transfer from landings under CAS technology to landings under PLM. These findings are consistent with Muller, McLaren, Appleby, & Rosalie (2015), where they used a temporal occlusion paradigm to establish that expert rugby players showed high perceptual skill in recognizing rugby situations, but that their skill did not transfer to related perceptual skills, such as recognizing baseball pitches.

Figure 6. Non-decision time (τ) results. Training increased non-decision time for experts using both technologies and for novices using PLM. For PLM, this is likely due to the need for relearning, or at least reinforcing, new perceptual patterns.

CONCLUSIONS AND FUTURE RESEARCH

The development of perceptual motor skills is an important part of military training and key components of those skills are implicitly learned perceptual patterns. Drift diffusion modeling of the component parts of those skills can provide insight into how those skills operate. Speed of processing reflects an important aspect of expertise, but experts—at least in this study—also use some of the time they gain to gather more information before making a decision. Experts also encode the situation faster.

However, expertise does not seem to transfer easily. Consistent with studies of other kinds of expertise transfer (e.g., Cheung & Bar, 2012; Muller, McLaren, Appleby, & Rosalie, 2015; Gauthier et al., 2002), participants who were expert in CAS but not expert in PLM showed a lower speed of processing and required more time to encode the situation under PLM, as compared with both themselves under CAS and with novices with the new technology. This likely was due to the need to unlearn the very skills and patterns that made them experts. Fortunately, experts caught up to novices with speed of processing and situation encoding after 10 or so simulator landings with the new technology (approximately 1 hour of instruction) —and in any case, landing performance was far superior under the PLM for everyone despite limited exposure to the technology.

In future research, it will be interesting to assess the effects of training on other kinds of learning. It is likely that different training regimens improve different DDM parameters. For example, two related hypotheses might be that deliberate practice has its primary effect on speed of processing as automaticity kicks in, but that exposure to a wide variety of conditions and stimuli has its primary effect on the decision threshold, since learning about more conditions could lead to a more complicated decision-making process. If these hypotheses are valid (or in fact if different kinds of training have differential effects of any sort on these components), then it will be possible to adapt training requirements to individual DDM strengths and weaknesses to specifically focus on the weakest components.

DDM may also provide a means to measure the components of adaptability. Although DDM doesn't measure adaptability directly, it can provide insights regarding the benefits of different types of cognitive training. In particular, DDM will provide more granular criterion measures than a standard grading rubric. For example, it can measure the
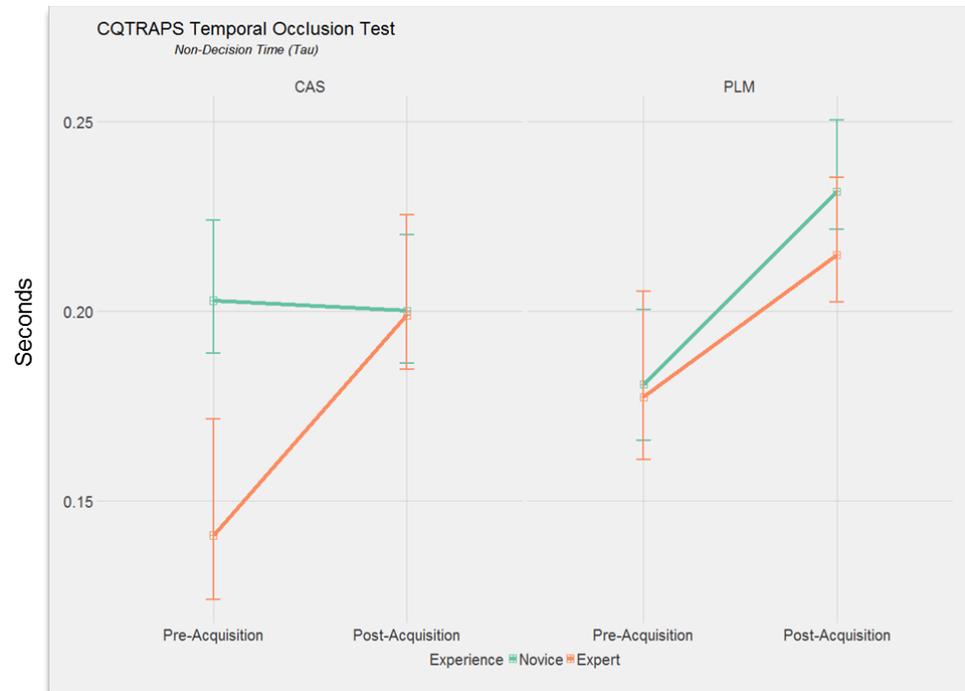
relative importance of speed of processing, the decision criterion, and non-decision time in adaptability, and thus can help focus adaptability training on the most important aspects for each individual trainee.

Beyond adaptive DDM training, it will be interesting to explore the effects of occlusion training for decision making under conditions of uncertainty and time stress. In sports, this training generally addresses perceptual or perceptual-motor skills, but occlusion training need not even have a perceptual component—all that is required is that the trainee is required to build a mental situational image, possibly from cues that do not directly map onto visual aspects of the situation. Progressive occlusion could be temporal, spatial, or simply cue-related, and trainees could be required to make accurate, speeded decisions. RTs and errors from those decisions could then be used to compute the components of DDM. Together, DDM and occlusion training can form the basis of accelerating the acquisition of expertise (Hoffman et al., 2013) by providing both the ability to identify cognitive components most in need of training and an approach for providing that training.

In short, DDM, especially when coupled with occlusion training, provides new opportunities for diagnosing and improving perceptual, motor, and cognitive skills, and for deepening our understanding of the nature of expertise. We think they have a very promising future and encourage others to use these techniques in their own research.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, J.R. (1995). A simple theory of complex cognition. *American Psychologist*, **51**(4), 255-365.

Anderson, J.R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Cheung, O.S. & Bar, M. (2012). Visual prediction and perceptual expertise. *International Journal of Psychophysiology*, **83**, 156-183.

Doob, J.L. (1953). *Stochastic Processes.* New York: Wiley and Sons.

Eastman, K., Stankiewicz, B., & Huk, A. (2007). Optimal weighting of speed and accuracy in a sequential decision-making task. *Journal of Vision*, **7**(9), 428.

Fadde, P. (2010). Look, Ma, no hands: Part-task training of perceptual cognitive skills to accelerate psychomotor expertise. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL.

Farrow, D, Abernathy, B., & Jackson, R.C. (2005). Probing expert anticipation with the temporal occlusion paradigm: experimental investigations of some methodological issues. *Motor Control*, **9**(3), 332-351.

Forstmann, B.U., Ratcliff, R., & Wagenmakers, E.-J. (2016). Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annual Review of Psychology*, 67, 641-666.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2002). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, **3**(2), 191–197.

Green, D.M., Swets J.A. (1966) *Signal Detection Theory and Psychophysics*. New York: Wiley.

Hoffman, R.R., Ward, P., Feltovich, P.J., DiBello, L., Fiore, S.M., & Andrews, D.H. (2013). *Accelerated Expertise: Training for High Proficiency in a Complex World.* Psychology Press.

Luce, R.D. (1986). *Response Times: Their Role in Inferring Elementary Mental Organization*. New York: Oxford University Press.

Kruschke, J. *Doing Bayesian Data Analysis, 2nd Ed.: A Tutorial with R, JAGS, and Stan* (2014). London: Academic Press.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, **44**, 314-324.

Muller, S., & Abernathy, B. (2012). Expert anticipatory skill in striking sports: A review and a model. *Research Quarterly for Exercise and Sport*, **83**(2), 175-187.

Muller, S., McLaren, M. Appleby, B., & Rosalie, S.M. (2015). Does expert perceptual anticipation transfer to a dissimilar domain? *Journal of Experimental Psychology: Human Perception and Performance*, **41**(3), 63-68.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20–22, Vienna, Austria.

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and negative responses. *Psychological Review*, **93**, 212-225.

Ratcliff, R. & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science,* **9**, 347-356.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and I@ in two-choice tasks. *Cognitive Psychology*, **60**, 127-157.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging*, **16**, 323-341.

Ratcliff, R. & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reactions times and parameter variability. *Psychonomic Bulletin & Review*, **9**, 438-481.

Ratcliff, R. & van Dongen, HPA. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin and Review*, **16**, 742-451.

Stacy, E.W., Beaubien, J.M., Wiggins, S.M., Walwanis, M., & Bolton, A. (2014). Using temporal occlusion to assess carrier landing skills. In *Proceedings of the 2014 Interservice/Industry Training, Simulation, and Education Conference*. Orlando, FL.

Stacy, W., Walwanis, M., Wiggins, S., & Bolton, A. (2013). Layered fidelity: an approach to characterizing training environments. *Proceedings of the Interservice/Industry Training, Simulation, and Education Conference*, Orlando, FL.

Strayer, D. & Kramer, A. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. **20**, 318-341.

Swets, J. A., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, **68**, 301-340.

Swets, J.A. (Ed.) (1964) *Signal detection and recognition by human observers*. New York: Wiley

van Ravenzwaaij, D., Donkin, C. & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review*, **24**(2), 547-556.

Wabersich, D. & Vanderkerckhove, J. (2104a). The RWiener package: An R package providing distribution functions for the Wiener diffusion model. *The R Journal*, **6**(1), 49-56.

Wabersich, D. & Vanderkerckhove, J. (2104b). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavioral Research*, **46**, 15-28.

Ward, P. Farrow, D. Harris, K.R., Williams, A.M., Eccles, D.W., & Ericsson, K.A. (2008). Training perceptual-cognitive skills: Can sport psychology research inform military decision training? *Military Psychology* **20**(Suppl 1), S71-S102.

White, C.N., Ratcliff, R., Vasey, M.W., & McKoon, G. (2010). Using diffusion models to understand clinical disorders. *Journal of Mathematical Psychology*, **54**, 39-52