



Michigan Technological University  
*Create the Future* Digital Commons @ Michigan Tech

---

Dissertations, Master's Theses and Master's  
Reports - Open

Dissertations, Master's Theses and Master's  
Reports

---

2013

## Assessing Decision Making Skill in Complex and Dynamic Environments Using Representative and Simulated Tasks

Patrick Karl Belling  
*Michigan Technological University*

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Cognitive Psychology Commons](#)

Copyright 2013 Patrick Karl Belling

---

### Recommended Citation

Belling, Patrick Karl, "Assessing Decision Making Skill in Complex and Dynamic Environments Using Representative and Simulated Tasks", Master's Thesis, Michigan Technological University, 2013.  
<https://doi.org/10.37099/mtu.dc.etds/668>

Follow this and additional works at: <https://digitalcommons.mtu.edu/etds>



Part of the [Cognitive Psychology Commons](#)

ASSESSING DECISION MAKING SKILL IN COMPLEX AND DYNAMIC  
ENVIRONMENTS USING REPRESENTATIVE AND SIMULATED TASKS

By

Patrick K. Belling

A THESIS

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

In Applied Cognitive Science and Human Factors

MICHIGAN TECHNOLOGICAL UNIVERSITY

2013

© 2013 Patrick K. Belling

This thesis has been approved in partial fulfillment of the requirements for the Degree of  
MASTER OF SCIENCE in Applied Cognitive Science and Human Factors.

Department of Cognitive and Learning Sciences

Thesis Advisor: *Dr. Paul Ward*

Committee Member: *Dr. Edward T. Cokely*

Committee Member: *Dr. Michele H. Miller*

Department Chair: *Dr. Brad Baltensperger*

To my mother and father, Jane and Craig, for their unconditional support and love.

## Table of Contents

List of Figures .....	7
List of Tables .....	8
Preface.....	10
Acknowledgements.....	11
Abstract.....	12
INTRODUCTION .....	14
Measuring Perceptual-Cognitive Skills in Sport: Differentiating between Skill Groups and Predicting Performance.....	16
Anticipation.....	16
Decision Making.....	19
Option Generation to Explain Expert Performance during Different Phases of Decision Making.....	22
Intervention phase.....	23
Assessment phase.....	27
Effects of Time Constraint on Performance and Strategy / Heuristic Use	31
Other Predictors of Skill in Sport.....	35
Domain-knowledge, recall, and recognition.....	36
Examining the relative contribution of domain-specific skills and domain-general abilities.....	42
Overview of the Current Research.....	47
EXPERIMENT 1 .....	50
Hypotheses about Performance during Assessment and Intervention.....	51
Hypotheses about Process during Assessment .....	52
Hypotheses about Process during Intervention.....	53
Methods.....	55
Participants.....	55
Materials .....	55
Procedure .....	59
Data Analysis .....	62
Coding of options on response sheet .....	62
Measures .....	63
Analysis.....	64
Results.....	68
Manipulation Check.....	68
Performance .....	70
Option Generation during Assessment .....	71

Option Generation during Intervention.....	74
Discussion .....	77
EXPERIMENT 2 .....	85
Hypotheses about Performance during Assessment and Intervention.....	86
Hypotheses about Process during Assessment .....	87
Hypotheses about Process during Intervention.....	88
Methods.....	90
Participants.....	90
Materials and Procedure .....	90
Data Analysis .....	91
Coding and Measures.....	91
Analysis.....	91
Results.....	92
Manipulation Check.....	92
Performance during Assessment and Intervention .....	94
Option Generation during Assessment .....	94
Option Generation during Intervention.....	96
Discussion .....	98
EXPERIMENT 3 .....	103
Hypotheses about Performance during Assessment and Intervention.....	104
Hypotheses about Process during Assessment .....	104
Hypotheses about Process during Intervention.....	105
Methods.....	106
Participants.....	106
Materials, Procedure, and Coding.....	107
Analysis.....	107
Results.....	108
Manipulation Check.....	108
Performance during Assessment and Intervention .....	109
Process during Assessment .....	110
Process during Intervention .....	112
Discussion .....	114
EXPERIMENT 4 .....	118
Methods.....	120
Participants.....	121
Materials .....	121

Procedure .....	125
Data Analysis .....	126
Coding and Measures.....	126
Analysis.....	126
Results.....	127
Discussion .....	128
EXPERIMENT 5 .....	130
Methods.....	132
Participants.....	132
Materials, Procedure, and Scoring.....	132
Analysis and Results .....	133
Discussion .....	136
EXPERIMENT 6 .....	137
Hypotheses about Performance on the OASSIS and Domain- General Cognition.....	139
Hypotheses about Performance on the OASSIS and Domain- Specific Knowledge.....	139
Methods.....	139
Participants.....	139
Materials and Scoring .....	140
Procedure .....	141
Analysis and Results .....	142
Performance on the OASSIS and Domain-General Cognition....	142
Performance on the OASSIS and Domain-Specific Knowledge.	142
Discussion .....	143
GENERAL DISCUSSION .....	144
References.....	152
Appendix: Supplemental Figures and Tables .....	164

## List of Figures

Figure 1 .....	57
The final frame of a video simulation (left) and the matching occlusion image (right) during the option generation paradigm.	
Figure 2 .....	58
An occlusion-image on-screen (left) and a matching response sheet with written and stamped options and ratings of likelihood and concern during assessment on the option generation paradigm.	
Figure 3 .....	123
The occlusion image for an action simulation of the OASSIS. Response options for pass, shoot, or dribble were displayed below the image.	
Figure 4 .....	123
The occlusion image for a direction simulation of the OASSIS. Response options for each direction on-screen were displayed below the image.	
Figure 5 .....	124
The occlusion image for a pass-recipient simulation of the OASSIS. Response options for each target player on-screen were displayed below the image.	
Figure 6 .....	136
Mediation table of option generation behavior, anticipation accuracy during assessment-phase trials of the option generation paradigm, and OASSIS performance.	
Supplemental figures in the appendix.....	164



## List of Tables

Table 1 .....	69
Anxiety and mental effort ratings during time constrained and non-time constrained assessment trials of Experiment 1.	
Table 2 .....	69
Anxiety and mental effort ratings during time constrained and non-time constrained intervention trials of Experiment 1.	
Table 3 .....	70
Mean (SD) frequency scores for anticipation accuracy during six time constrained and six non-time constrained assessment trials and selection of the criterion best option during six time constrained and six non-time constrained intervention trials, main effects of skill and time constraint, and interaction effects of skill by time constraint on performance during assessment and intervention trials of Experiment 1.	
Table 4 .....	72
Means (SD) number of options generated per trial and interaction effects of skill by information type on number of options generated during assessment trials under each time constraint condition of Experiment 1.	
Table 5 .....	75
Means (SD) number and type of options generated per trial and simple interaction effects of skill by information type on number of options generate during intervention trials under each time constraint condition of Experiment 1.	
Table 6 .....	93
Anxiety and mental effort ratings during time constrained and non-time constrained assessment trials of Experiment 2.	
Table 7 .....	93
Anxiety and mental effort ratings during time constrained and non-time constrained intervention trials of Experiment 2.	

Table 8 .....	94
Mean (SD) frequency (i.e., proportion) scores for anticipation accuracy during assessment and selection of the criterion best option during intervention and effects of time constraint for Experiment 2.	
Table 9 .....	95
Mean (SD) number and type of options generated, main effect of time, and interaction effect of Time x Information Type during assessment trials of Experiment 2.	
Table 10 .....	96
Mean (SD) number and type of options generated, main effect of time, and interaction effect of time by information type during intervention trials of Experiment 2.	
Table 11 .....	111
Mean (SD) number and type of options generated by the high- and low-skill groups during assessment-phase trials of Experiment 3.	
Table 12 .....	113
Mean (SD) number and type of options generated by the high- and low-skill groups during intervention-phase trials of Experiment 3.	
Supplemental tables in the appendix.....	172

## **Preface**

Parts of the research presented in this thesis have also been submitted for publication elsewhere. For full reference information, please see the reference list. The data reported on in Experiment 1 was also presented in Belling and Ward (2012). The data reported on in Experiment 2 was also presented in Belling, Suss, and Ward (2013; under review B). The data reported on in Experiment 3, Experiment 5, and Experiment 6 was also presented in Belling, Suss, and Ward (under review A). The data in Experiment 4 was presented in Belling and Ward (in press). Across all experiments, Patrick Belling collected and analyzed the data, and also wrote drafts of the findings. Dr. Paul Ward assisted with the analysis of data when needed and with the writing. Dr. Joel Suss assisted with the writing in the above experiments where he is listed as an author.

## **Acknowledgements**

I would like to acknowledge the U.S. Soccer Federation and the North American Society for the Psychology of Sport and Physical Activity (NASPSPA) Graduate Student Research Grant for their financial support of portions of the research presented in this manuscript.

Next, I would like to thank Department of Cognitive and Learning Sciences at Michigan Technological University for their support of my research.

I would also like to thank my committee members, Dr. Edward Cokely and Dr. Michele Miller, for their time and effort in reviewing this research.

Last, but not least, I would like to thank my final committee member and graduate advisor, Dr. Paul Ward, for his time, effort, knowledge, and patience while guiding me through not only the research presented in this manuscript, but also the academic journey more broadly.

## **Abstract**

Traditional decision making research has often focused on one's ability to choose from a set of prefixed options, ignoring the process by which decision makers generate courses of action (i.e., options) in-situ (Klein, 1993). In complex and dynamic domains, this option generation process is particularly critical to understanding how successful decisions are made (Zsombok & Klein, 1997). When generating response options for oneself to pursue (i.e., during the intervention-phase of decision making) previous research has supported quick and intuitive heuristics, such as the Take-The-First heuristic (TTF; Johnson & Raab, 2003). When generating predictive options for others in the environment (i.e., during the assessment-phase of decision making), previous research has supported the situational-model-building process described by Long Term Working Memory theory (LTWM; see Ward, Ericsson, & Williams, 2013). In the first three experiments, the claims of TTF and LTWM are tested during assessment- and intervention-phase tasks in soccer. To test what other environmental constraints may dictate the use of these cognitive mechanisms, the claims of these models are also tested in the presence and absence of time pressure. In addition to understanding the option generation process, it is important that researchers in complex and dynamic domains also develop tools that can be used by 'real-world' professionals. For this reason, three more experiments were conducted to evaluate the effectiveness of a new online assessment of perceptual-cognitive skill in soccer. This test differentiated between skill groups and predicted performance on a previously established test and predicted option generation

behavior. The test also outperformed domain-general cognitive tests, but not a domain-specific knowledge test when predicting skill group membership. Implications for theory and training, and future directions for the development of applied tools are discussed.

# Assessing Decision Making Skill in Complex and Dynamic Environments

## Using Representative and Simulated Tasks

### **Introduction**

On July 3<sup>rd</sup>, 1988, the Iran Air Flight 655, a civilian airliner, was attacked and destroyed by the USS Vincennes. All 290 passengers aboard were killed and relations with Iran have declined since. Because of poor expert decision making by what society deemed expert military personnel, severe and permanent consequences were set in place. While we cannot hope to correct the judgment and decision making of the past, we can attempt to prevent these types of decision errors in the future. Research that increases our understanding of, and creates tools to assess, the cognitive mechanisms that facilitate expert decision making in stressful environments such as this will be a critical step towards prevention.

There are several objectives in the current research: First, I provide an overview of the research that has measured skill-based differences in perceptual-cognitive skill that could be considered two ‘phases’ of the decision making process: (i) Making decisions about the courses of action that others in the environment will take next and/or how the situation will unfold (henceforth, the ‘assessment’ phase); and (ii) making decisions about the personal course(s) of action that the participant themselves will take next (henceforth, the ‘intervention’ phase).

Second, I review the claims of current theoretical models that provide different explanations about the (option generation) strategies employed by skilled individuals during these different phases of decision making. Performance is measured and tests of

these claims are conducted in the complex domain of soccer. Sport, and especially soccer, is one of the original domains in which these claims were made and it provides a useful model environment for other complex domains—it is a dynamic domain in which individuals are required to act upon partial or incomplete information, work both independently and as a team in an effective manner, and operate under stressful circumstances (see Ward, Farrow, Harris, Williams, Eccles, & Ericsson, 2008).

Third, I review the literature on the effects of one of the most common stressors found in complex and dynamic environments, time pressure, on intervention- and assessment-based decisions. Specifically, I examine whether the strategies and heuristics proposed to explain superior decision making under non-time constrained conditions are more or less consistent with those employed under time pressure.

Lastly, I develop a *short* test that is able to differentiate between skill levels, widely accessible, and easily implemented. Several tests of perceptual-cognitive skill focusing on assessment and/or intervention decisions already exist. However, to date none of these are widely accessible, and none have been used to predict performance and process (i.e., option generation strategies) on more in-depth tests, such as those that will be presented in Experiments 1-3. The focus of Experiments 4-6 is to determine the value of such a predictive test in comparison to other potential domain-specific and domain-general predictors of soccer skill.



## **Measuring Perceptual-Cognitive Skills in Sport: Differentiating between Skill Groups and Predicting Performance**

To begin, I will review the different types of tests that have been developed to measure constructs that are synonymous with the assessment and intervention phases, namely *anticipation* and *decision making* skills in sport. I will examine the methods employed and discuss if and how these have been validated.

**Anticipation.** Several methods have been used to measure anticipation, including spatial and temporal occlusion and response time paradigms (e.g., for a review see Williams & Ward, 2007). One of the most widely used tests to measure the perceptual-cognitive skill of anticipation is the temporal occlusion paradigm, originally introduced by Haskins (1965)—where dynamic information (e.g., video footage of game play) is presented up until a crucial moment (e.g., decision point, immediately prior to an action, such as 300ms prior to foot-to-ball contact) and then unexpectedly occluded from view. In studies that have employed this method, researchers have either manipulated the moment at which the occlusion occurs on-screen using video simulations (e.g., Abernethy, 1990; Abernethy & Russell, 1987) or used helmets/glasses in real-game-type conditions that occlude vision in-situ (cf. Starkes, Edwards, Dissanayake, & Dunn, 1995; Rosalie & Muller, 2013; for a review, see Farrow, Abernethy, & Jackson, 2005). At the moment of occlusion, participants are required to anticipate what happens next. This approach has been employed to demonstrate skill-based differences in anticipation accuracy in many sports, including squash (e.g., Abernethy, 1990), baseball (e.g.,

Burroughs, 1984), badminton (Abernethy & Russell, 1987), and soccer (Williams & Davids, 1995). Researchers have also demonstrated that training based on temporal occlusion can be effective and transfer to the ‘real-world’ performance in sports such as baseball (Fadde, 2006), softball (Gabbett, Rubinoff, Thorburn, & Farrow, 2009) and tennis (e.g., Williams, Ward, Knowles, & Smeeton, 2002).

When using the temporal occlusion paradigm, the research has demonstrated that, in general, skilled decision makers are able to anticipate the actions of others sooner than less-skilled decision makers by making use of environmental information (i.e., cues) that is made available in the early stages of the development of an action of an individual (e.g., the direction of the grounded foot, hips, torso, and head prior to striking the ball during a shot in soccer) (i.e., Abernethy & Russell, 1987; Goulet, Bard, & Fleury, 1989; Jones & Miles, 1978; Wright, Pleasants, & Gomez-Meza, 1990) or pattern of play involving multiple teammates (e.g., the movements and relative position of teammates ‘off-the-ball’ on a soccer field prior to the player in possession striking the ball) (e.g., Williams & Davids, 1995, Ward & Williams, 2003). Accordingly, occlusion conditions that limit access to information presented early (as opposed to late) in a dynamic action/pattern sequence often yield the highest skill-based differences and, hence, provide known-groups validity.

An early study that employed temporal occlusion investigated expert-novice differences in anticipation in badminton. Abernethy and Russell (1987) occluded video footage of an opposing badminton player at different time points around

racket/shuttlecock contact and demonstrated that expert players anticipated the outcome more accurately and were able to use information that occurred early in the action sequence (i.e., information available prior to contact) more effectively than novices. Further analysis of eye gaze data revealed that novices fixated the movement of the racket while anticipating, whereas experts utilized the movement of both the racket and the arm. Temporal occlusion, in this instance, not only revealed skill-based differences in perceptual-cognitive skill, but also provided insight into when important information cues were available in the environment and when these skills could be used by experts to gain a tactical advantage.

Similarly, Abernethy (1990) temporally occluded video clips at varying points during the swing stroke of an on-screen squash player. Participants were required to anticipate the direction and power of the occluded shot. Across all of the film conditions, experts were significantly more accurate at anticipating the event outcome than novice counterparts, and made use of information available early in the action sequence that novices were unable detect/use.

A more recent example in the literature is provided by Rosalie and Müller (2013). Expert, near-expert, and novice karate athletes were instructed to block attacks from an opponent while having their vision occluded using PLATO glasses. Vision was occluded prior to the opponent's movement (condition 1), after initial head movement of the opponent (condition 2), and after the initiation of attacking movement (condition 3). Lastly, a control condition was created with no occlusion (i.e., participants blocked

attacks from an opponent without any visual occlusion). Analysis of the skill groups' ability to anticipate the attack and block successfully revealed that experts performed significantly above chance regardless of occlusion condition. Near-experts, however, only performed significantly above chance after the initiation of an attacking movement (i.e., condition 3) and during the control condition (i.e., no occlusion). Lastly, novices only performed above chance during the control condition. The authors not only provided support for the finding that expert athletes rely on information earlier in the action sequence during karate, but also differed from near-expert and novice karate athletes in their ability to perceive and utilize visual information from a completely static opponent to anticipate attacks (Rosalie & Muller, 2013; cf. Borgeaud & Abernethy, 1987).

In summary, temporal occlusion has been used widely to differentiate between skill groups in their ability to anticipate how the situation will unfold. Laboratory tests have typically used video or live-action stimuli that present information up until a critical decision moment and then require participants to anticipate the potential consequences of the dynamic situation. In general, laboratory studies employing this technique have found that domain experts are able to anticipate more accurately than domain novices. Furthermore studies have demonstrated that experts can anticipate outcomes correctly sooner than novices.

**Decision Making.** While anticipation is an important component to performance within complex and dynamic domains, the ability to quickly act on that information is equally important. In fact, many individuals have examined the skill of athletes at

anticipating the actions of their opponent by measuring the course(s) of action taken in response to their opponent (e.g. Savelsbergh, Williams, van der Kamp, & Ward, 2002). I will now review how decision making in sport has been assessed in the lab throughout many years.

In an initial study that used decision making to differentiate between skill groups, Helsen and Pauwels (1988) used a live action task to assess the decision making of soccer players. Participants stood in front of a life-size projector screen and had a ball placed in front of them. When the video indicated that the ball was being passed to them, the players were instructed to make a decision, and kick the ball as they would in the game. Rather than finding significant differences between skill groups in decision making accuracy, Helsen and Pauwels (1988) observed significant skill-based differences in the speed with which correct decisions were made—when making the correct decision (e.g., shoot at goal, pass to correct player)—expert players made decisions more quickly.

In another study of skill-based differences in decision making, Ripoll, Kerlizin, Stein, and Reine (1995) examined the decision making and visual search of expert, intermediate, and novice boxers. Participants viewed stimuli depicting an opponent boxer that attacked in one of two experimental conditions—simple and complex situations. During simple situations, participants were required to respond to only one type of boxing move (e.g., attacks, openings, or feints). During complex situations, participants were required to respond to multiple types of boxing moves. Participants responded to the dynamic scenarios using a joystick. Participants were trained to respond using a specific

joystick motions (e.g., up, down) to signify a pre-learned response action (e.g., block/defend head, block/defend chest). Analysis of the skill-groups' performance revealed that skill-based differences only existed during complex situations (i.e., when responding to multiple types of attack move). Experts were more accurate in their responses than novices, but reaction time did not vary across groups.

In a similar study, Savelsbergh, Williams, Van Der Kamp, and Ward (2002) used a joystick-based response to skill-based differences of soccer goal keepers when facing a penalty kick. Specifically, expert and novice goalkeepers were instructed to move a joystick to respond to a series of penalty kick video simulations. Movement of the joystick indicated which direction they would dive (e.g., down and left, up and right, middle and left). Expert goalkeepers responded more accurately than novice goalkeepers. However, they waited longer than novices to make a response but made fewer corrective movements. These data suggest that while experts used late information to confirm their early anticipations (as suggested by the temporal occlusion research), novices used late information to correct erroneous decisions that resulted from their inability to correctly interpret early cues.

More recently, Gorman, Abernethy, and Farrow (2013) investigated decision making in low- and high-skill basketball players. Players were presented with either moving scenarios (i.e., brief video clips of 5-on-5 basketball play) or static scenarios (i.e., the last picture frame from a moving scenario video clip) and were required to indicate the action that they should perform next as the ball carrier. Actions included pass, dribble

right, dribble left, or shoot. An expert panel generated all of the acceptable actions for each trial. Participants were awarded points in the decision making task if the action that they selected matched an action that was present on the expert list. Significant skill differences were found in the decision making task in favor of the high-skill players. Expert players made more accurate decisions than novice players. There was also a main effect for display type, which indicated that more accurate decisions were made using still images. A skill by display type interaction was not observed.

While measuring anticipation and decision making in the lab in order to differentiate between skill groups can be useful, particularly in identifying the types of cognition that support expert performance, it does not necessarily offer a process-level description of the aforementioned phases of expert decision making. The next section will review the strategies and heuristics that have been proposed to explain, from a cognitive standpoint, how expert performance occurs. More specifically, I will review the literature that focuses on how experts generate courses of action (i.e., options) in-situ.

### **Option Generation to Explain Expert Performance during Different Phases of Decision Making**

Historically, traditional decision making research has investigated strategies for choosing between a fixed set of options presented to a participant in a task (cf. Zsombok & Klein, 1997). In early decision making research, the focus was often on the ability to make decisions economically, such as by expected utility (von Neumann & Morgenstern,

1947). Likewise, considerable work has focused on situations in which participants have deviated from such optimal strategies (e.g., heuristics and biases; Tversky & Kahneman, 1975). Unfortunately, this line of research has often ignored the process by which decision makers in naturalistic contexts actually generate the alternatives from which to choose (i.e., option generation). This process is critical in real-world decision making (Klein, 1993). Option generation has been used to provide a process-based explanation of skilled anticipation (Ward et al., 2013) and skilled decision making (Johnson & Raab, 2003; Raab & Johnson, 2007) within sport domains. Before discussing this research, to add some context to this research, I will first provide an overview of relevant process-oriented research in other complex and dynamic domains. I begin with a focus on option generation during the intervention phase of decision making since the majority of research has been conducted in this area.

**Intervention phase.** The Recognition Primed-Decision (RPD) model of rapid decision making (Klein, 1989; 1993) is an influential and often cited explanation of how experts generate courses of action in-situ. According to RPD (see Klein 1989; 1993), skilled decision makers operating within their domain of expertise engage in two mental processes when making decisions in complex and dynamic environments. First, experts engage in a pattern recognition process in which the situation is recognized and a ‘typical’ response for this type of situation is selected. If necessary, experts then mentally simulate the selected reaction to check for its plausibility. If the response is deemed to be appropriate, the expert will execute the necessary action. If deemed inappropriate after



mental simulation, the expert will generate the next most typical response and re-engage in mental simulation. This process cycles until an appropriate action is selected. For these reasons, RPD predicts that expert decision makers will generate very few (i.e., 1 or 2) options and higher skill will be associated with fewer options generated—because they are more likely to generate appropriate responses based on their experiences to date. Klein (1993) provided support for RPD by interviewing and obtaining protocols from urban fireground commanders. Based on critical decision method interviews (see Klein, Calderwood, & Macgregor, 1989) performed by Klein (1993), the fireground commanders were not ‘weighing alternatives’ or ‘assessing probabilities’ as is often cited in traditional decision making research (Zsombok & Klein, 1997). Instead, they claimed to simply be *reacting* to situations based on relevant experience and training.

In an investigation of skilled decision making, Klein (1995) examined the option generation behavior of skilled chess players. Participants were presented with four different configurations of pieces on a chess board and were required to generate the next move. During the move selection task, participants were instructed to verbalize every potential move they were considering, regardless of its perceived strength. Of 124 legal moves that could be made in a given scenario, only 20 were subsequently rated as ‘acceptable’ moves by an independent panel of grandmaster (i.e., expert) chess players. However, 41 of the 64 moves that participants generated as their *first* option were one of those 20 rated as acceptable—suggesting the first option was an acceptable option more often than would be expected by chance. This provided support for Klein’s (1995)

primary hypothesis—skilled decision makers were able to generate satisfactory options as their first option considered.

Thus far, Klein's work (1989; 1993; 1995) demonstrated that skilled participants tended to generate few options and better ones first during the intervention-phase of decision making—a strategy consistent with the fast and frugal philosophy of 'less is more'.

Johnson and Raab (2003) also found support for a 'less-is-more' phenomenon regarding option generation during the intervention phase of decision making. In an investigation of option generation behavior in handball, these authors expanded upon the predictions of RPD and proposed the Take-The-First heuristic (TTF) to explain successful decision making. The TTF is a fast and frugal heuristic—i.e., a heuristic that offers an explanation of human behavior and reasoning using limited information and time (Gigerenzer & Todd, 1999). The TTF fits the fast and frugal claim quite well based on its ability to explain how temporally-constrained decision making occurs. In response to complexity and time demands that put incredible pressure on the decision maker, consistent with RPD, TTF suggests that skilled decision makers select the first option generated. The first option is often 'satisficing' (i.e., good enough or a workable solution) and is selected without much deliberation (see below for an explanation of the underlying mechanism). In their initial study, Johnson and Raab (2003) examined the decision strategy of handball players of varying skill levels. Participants were shown videos of handball play which froze immediately prior to the player with the ball making a decision

about what to do next (e.g., shoot, pass). The still image from video was left on-screen for participants. Participants were then asked to generate the options that they would consider performing if they were the player with the ball on-screen. Their data indicated that skilled and experienced decision makers in a dynamic environment generate relatively few (e.g., 2-3) intervention options (rather than only one; cf. Klein, 1993). In addition, they provided further support that options were generated serially (see Klein, 1993), and a sacrificing option was often generated first.

As a means to explain how superior decision makers generate good options first, proponents of TTF suggest that the initial option is generated based on association with the environmental structure. As activation spreads, other less relevant options are generated and the quality of subsequent options generated relative to the first option decreases. As the number of options increases, the quality of the final selected intervention option decreases because the risk of selecting a poorer option increases. As a consequence, Johnson and Raab (2003) predicted and observed an inverse relationship between the number of options generated and final decision quality. In a subsequent study of handball players, Raab and Johnson (2007) also demonstrated that a positive correlation existed between level of expertise and TTF. In other words, as one acquires more skill, they follow the trends predicted by TTF more closely. In summary, proponents of both TTF and RPD suggest that skilled decision makers often use simple heuristics during the intervention phase of decision making in complex and dynamic environments.

**Assessment phase.** The RPD and TTF offer explanations for skilled decision making, but focus on the intervention phase of the decision process (i.e., generating courses of action for oneself to pursue). Research has also been conducted on option generation during the assessment phase of the decision process. Klein and Peio (1989) demonstrated that the principles of RPD may apply also when predicting the outcome of a situation. Klein and Peio (1989) studied the ability of high- and low-skill chess players to generate predictions about their opponents' moves. They found that proficient players were more accurate at predicting the next move, more likely to generate the next (i.e., correct) move as the first one generated in a sequence of alternatives, and made fewer guesses in total about the predicted move.

On the other hand, some researchers have not observed a 'less-is-more' phenomenon during the assessment phase of decision making. In a review of research on expert performance Ericsson and Kintsch (1995) suggested that experts build a situational model that integrates stored knowledge with new environmental information *on-the-fly*, producing an updated situational representation. These authors termed the ability to index and encode information in an accessible form via such a mechanism as Long Term Working Memory (LTWM) skill. Such a mechanism allows experts to not only maintain direct access to task-relevant information but to accurately anticipate situational consequences and access appropriate decision alternatives. Based on Ericsson and Kintsch's (1995) LTWM theory, Ward, Ericsson and Williams (2013) predicted that skilled decision makers would generate more task-relevant (i.e., 'high quality') options

than less skilled decision makers during an anticipation task. The generation of more task-relevant options was predicted to be the result of a more sophisticated situational model developed through extensive task-relevant experience, such as in deliberate practice (see Ericsson, Krampe, & Tesch-Römer, 1993). Unlike TTF, Ward et al. (2013) did not make predictions regarding the relationship between *total* options generated and final decision quality/skill. Instead they made predictions about the relationships between skill, decision quality, task-relevant options, and task-irrelevant options—where ‘task-relevant’ options were considered all of the options judged to be of high quality by an expert panel (i.e., those that an expert should have heeded and considered to have some degree of threat). All options generated by participants that were not considered to be of high quality were deemed ‘task-irrelevant’. Similar to TTF, Ward et al.’s predictions based on LTWM indicate that experts generate fewer task-irrelevant options than novices. However, counter to TTF, Ward et al. predicted a positive relationship between the number of task-relevant (i.e., high quality) options that can be generated and final decision quality/level of expertise.

In a test of their predictions derived from LTWM, Ward et al. (2013) required participants to watch dynamic video scenes of soccer play which either froze on screen (Exp. 1) and/or were occluded (Exp. 2 & 3) at critical decision moment. From the perspective of a defensive player, participants were then asked to generate options that were heeded when the play was frozen/occluded and available to the player with the ball, and to simultaneously anticipate the outcome of the play. High-skilled participants

outperformed low-skilled participants in terms of anticipation accuracy, the number of task-relevant options generated, and the prioritization of those options according to the threat posed to the defense. The number of task-relevant options generated was positively correlated with anticipation accuracy, both of which differentiated skill groups in favor of the high-skill participants. Skill effects were stronger in a non-cued condition, where all perceptual information was removed from the screen at the critical decision moment. This forced participants to generate anticipation options without being allowed to maintain perceptual access to information on screen (e.g., the location of players). Ward et al. (2013) described this finding as the result of maintained access to a well-developed situational model by skilled participants. As perceptual information is removed, reliance upon the situational model is increased, resulting in an even stronger advantage for high-skilled individuals with more developed mental representations of the dynamic patterns of game play in the natural ecology.

In addition to the use of an occlusion (rather than freeze-frame/cued) condition, Ward et al. (2013), pointed out methodological differences between their research and work done by proponents of TTF (Johnson & Raab, 2003; Raab & Johnson, 2007) and RPD (Klein, 1989; 1993). Firstly, Ward et al. (2013) emphasized the importance of the distinction between task-relevant (i.e., high quality) and task-irrelevant (i.e., low quality) options—because the associated situational model was expected to contain a detailed, accurate and up-to-date representation of relevant information currently available in the environment. Instead of generating hypotheses regarding the total number of options

generated, Ward et al. (2013) demonstrated that performance on an option generation paradigm was positively and negatively related to the generation of task-relevant and task-irrelevant options, respectively. Secondly, Ward et al. (2013) drew attention to important theoretical distinctions that may dictate the use of the various proposed cognitive mechanisms facilitating expert performance. Although previous research has not examined how experts utilize these strategies in different situations, preliminary data suggests that skilled decision makers might employ different strategies under different conditions and in the different phases of decision making (Belling & Ward, 2012; Ward, Suss, Eccles, Williams, & Harris, 2011). In general, when selecting a course of action for oneself (i.e., during intervention) research supports the use of a simple heuristic such as TTF (Johnson & Raab, 2003, cf. Ward et al., 2011). Alternatively, when selecting a course of action for others in the environment and/or anticipating the outcome of a situation (i.e., during assessment), research in dynamic domains supports the use of LTWM mechanisms (Ward et al., 2013). Likewise, when operating under time pressure, one might hypothesize that participants would make greater use of an RPD or TTF-type strategy (which are empirically consistent in regards to option generation)—even in the assessment phase of decision making—simply because of the limited amount of time available to generate options. In the next section, I review past research, where available, that has examined the effect of time constraint on anticipation and decision making and on option generation behavior.

## **Effects of Time Constraint on Performance and Strategy / Heuristic Use**

According to Herbert Simon (1969), "Human beings, viewed as behaving systems, are quite simple. The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves." (p. 53).

Zsombok and Klein (1997) describe the constraints that exist on the naturalistic domains, in terms of environmental complexity, that are of interest to the current research. Such domains include: (i) ill structured problems, (ii) uncertain and dynamic environments, (iii) shifting, ill-defined, or competing goals, (iv) action and feedback loops, (v) high stakes, (vi) multiple players, (vii) organizational goals, and (viii) time stress.

The domain of sport, and more specifically soccer, is highly representative of a range of naturalistic human performance domains with respect to each of these constraints. While individuals in sport and other complex domains often operate under intense time pressure (e.g., handball, Johnson & Raab, 2003; soccer, Ward et al., 2013; law enforcement, Ward et al., 2011; fire fighting, Klein, 1993), rather surprisingly, research that manipulates the effect of time constraint on performance and/or process (e.g., option generation) is limited.

In choice-reaction tasks both speed and accuracy can be important variables and the body of research to date suggests that the quality of performance is mediated by a speed-accuracy trade off. Proctor and Van Zandt (2008) describe the speed-accuracy trade-off (see also Osman, Lou, Muller-Gethmann, Rinkenauer, Mattes, & Ulrich, 2000) in choice reaction tasks as follows:



...if accuracy is of no concern, you could make any response you wanted whenever you detected the onset of a stimulus. You would simply be guessing at the appropriate response, which you can then make very quickly, but your accuracy would be no better than chance. Alternatively, you could wait until you were sure about the identity of the stimulus and its associated response. Your responses would be much slower, but your accuracy would be perfect. (p. 343).

While research on the speed-accuracy trade-off has been conducted using psychomotor tasks primarily, it suggests that participants' level of accuracy on these tasks will decrease with the implementation of time constraint. However, some evidence suggests that skilled performers are not as affected by time pressure—especially in knowledge-based or cognitively oriented tasks—and therefore do not experience a degradation in accuracy (Calderwood, Klein, & Crandall, 1988; Gobet & Simon, 1996a).

Researchers have constrained the time available to respond in studies of expertise in chess. Calderwood et al. (1988) demonstrated that the quality of moves by master chess players was unaffected by time constraint. Moves made in time-pressured games (i.e., approximately 6s per move) were not of lower quality than those in standard tournament games (i.e., approximately 135s per move).

Another method for increasing time pressure is playing in simultaneous games. Playing in multiple chess games simultaneously considerably reduces the time available to select the next move in each individual game. Gobet and Simon (1996a) compared the performance of Gary Kasparov, a highly elite chess player, during standard tournament

games and during 4-8 simultaneous games. Standard tournament games allow players three minutes to make each move. During simultaneous matches, Kasparov was therefore limited to 3 minutes in making all 4-8 of his moves, whereas his opponents were playing under standard time limits (i.e., 1 move per 3 minutes). Kasparov's performance did not degrade significantly during simultaneous matches. The authors concluded that grandmaster (i.e., expert) chess players were relatively unaffected by time pressure.

Contrary to this finding, Chabris and Hearst (2003) demonstrated that the performance of grandmaster chess players degraded significantly while under time pressure. The authors required these players to compete against one another during rapid and standard games and recorded the number of blunders (i.e., bad moves). Grandmaster players made significantly more blunders while playing against other grandmasters during rapid games (i.e., 25-minute time limit for the entire game, averaging less than 1 minute per move) as opposed to standard games (i.e., averaging 3 minutes per move).

Other studies have also shown that time pressure can affect skilled performance. For instance, in a study of humans playing computer-generated chess moves, Lassiter (2000) demonstrated when the available amount of time to select a move decreases, human chess experts are less and less likely to succeed against computer counterparts set to the same level of skill. When computers and humans of the same difficulty/rating level faced off, computers had a significant advantage when the chess game was limited to 25 minutes (i.e., 30 seconds per move on average). This advantage was doubled when

the game was limited to 5 minutes (i.e., 6 seconds per move on average). Again, this suggests skilled performers can be affected by time constraint.

Based on this line of research, the effect of time constraint has yielded somewhat controversial results, even in traditionally cognitively oriented tasks (e.g., chess).

Considerable research has demonstrated that time pressure can affect performance (Proctor & Van Zandt, 2008; Chabris & Hearst, 2003; Lassiter, 2000); though some have suggested that expert performance is not as affected (Calderwood et al., 1988; Gobet & Simon, 1996a). While considerable work has been done in an attempt to explain expert performance in temporally-constrained domains (cf. Klein, Oransu, Calderwood, & Zsombok, 1993; Klein, 1989, 1993; Joslyn & Hunt, 1998; Johnson & Raab, 2003; Raab & Johnson, 2007; Ward et al., 2013), research that actually *manipulates* time constraint within a specific representative task is very limited—despite the apparent prevalence of temporally-related factors within these domains. In one study, Raab and Johnson (2007) significantly limited the exposure time to the stimulus—the final frame of action left on screen as a still image—test (i.e., 6 seconds) relative to previous research (i.e., 45 seconds; Johnson & Raab, 2003). While the time available to generate options in both studies was unlimited, the reduction in exposure time to the stimulus did not appear to affect performance and/or process (i.e., option generation behavior). By actually manipulating time constraint within a given representative task, I am able to test the claims of multiple models of decision making and uncover when (i.e., under which conditions) each model is capable of predicting skilled performance.

In the current research, I investigate the effect of time constraint on option generation behavior and skilled performance in sport. This is achieved by presenting video footage of dynamic patterns of sports play and requiring participants to generate relevant options about what may have occurred had the video not ended (e.g., Ward et al., 2013), and what they would have done at that point (e.g., Johnson & Raab, 2003; Raab & Johnson, 2007). In addition to the methods employed by previous researchers in dynamic (e.g., sport) domains, I manipulate the time to respond during trials, similar to the work done in static domains (e.g., chess). Specific hypotheses regarding the effect of time constraint are provided within experiment descriptions.

In addition to the manipulation of decision perspective (e.g., assessment and intervention phases) and time constraint to inform theory on skilled decision making, I also leverage the information provided by previous research in order to apply theory within a domain using a new tool. In the design of this applied tool, I also consider what other constructs may predict skill. The literature on this topic is reviewed in the following section.

### **Other Predictors of Skill in Sport**

A primary goal of this research is to examine the validity of shorter tests of perceptual-cognitive skill (i.e., temporal occlusion test of anticipation). More specifically, my intent is to examine the predictive validity of such a test with longer, more difficultly implemented tests of perceptual-cognitive skill that measure both performance and

process (i.e., option-generation paradigm). Assuming predictive validity can be demonstrated, my intent is to examine the extent to which such a short test can predict skill level in comparison to other tests that may also have some predictive power. To provide some context for the choice of tests that will be used as predictors of skill level—alongside a short test of perceptual-cognitive skill—I now provide a review of other constructs have either been used to differentiate between skill groups and/or have demonstrated convergent validity with measures of anticipation. In the first instance, I examine measures of domain-specific knowledge and memory.

**Domain-knowledge, recall, and recognition.** Knowledge has been used to predict skill in the past. In a well-known description of knowledge acquisition and structure, Anderson (1982) made the distinction between declarative knowledge and procedural knowledge. Declarative knowledge refers to ‘what’ to do. This includes all relevant factual information. For example in sport, this might include the rules, playing surfaces, and tactics, among other information. Procedural knowledge refers to ‘how’ to do. This refers to specific skills and abilities developed. For example in sport, this might include knowledge of how to dribble a basketball, how to serve a tennis ball, or how to shoot a soccer ball.

French and Thomas (1987) required young (e.g., 8-12 years old) basketball players of varying (e.g., high/low) skill to complete a declarative and procedural knowledge test. To test declarative knowledge, the players completed a basketball knowledge test. To test procedural knowledge, the players also completed basketball

shooting and dribbling tests. Lastly, French and Thomas (1987) also recorded each player's ball control, decision making accuracy, and ability to execute their selected decisions. The high-skill group scored higher on the basketball knowledge test, shooting, and dribbling tests, and also performed better in games (i.e., controlled the basketball better, made more accurate decisions, and executed selected decisions better).

Furthermore, decision making accuracy in-game was related to performance on the basketball knowledge test. These results suggest declarative and procedural knowledge both contribute to expertise and can be used to discriminate between skill levels in sport.

In a follow up study, French and Thomas (1987) observed how knowledge and performance developed over the course of a season. While the dribbling and shooting skill tests remained relatively constant, the declarative knowledge test improved—suggesting that young players may gain declarative knowledge faster than they are able to improve motor skills in a sport. Moreover, the increase in declarative knowledge throughout the season predicted decision making ability at the end of the season, providing further support for the importance of domain-specific declarative knowledge in skill development. McPherson and French (1991) also demonstrated that domain-specific knowledge increases as a result of skill-based- and/or strategy-based instruction in tennis. In addition to domain-specific knowledge differentiating skill groups, domain-specific memory has been used by a number of researchers.

Recall accuracy of domain-specific information—an indirect test of how knowledge is organized and represented in memory—has also been shown to

discriminate between skill groups. In the seminal work by Chase and Simon (1973), multiple skill groups of chess players were shown a chess board that contained a configuration of pieces. After exposure, participants were required to re-create the chess board by recalling the locations of each piece on the board. The initial configurations of pieces were structured (i.e., pieces were placed on the board in patterns that occur in real game situations) or unstructured (i.e., pieces were randomly placed on the board). Expert chess players demonstrated significantly higher recall of structured (i.e., representative) formations of pieces than novice players, suggesting superior domain-specific memory. However, no significant skill differences were observed for unstructured configurations (i.e., domain-general memory). The authors suggested experts are able to encode, store, and access relevant ‘chunks’ (i.e., patterns of pieces) that they have encountered through extensive experience.

The superiority of experts in recall of relevant ‘chunks’ of information has been extended from the seminal work in chess (see Chase & Simon, 1973) to dynamic domains, such as various sports (Allard, Graham, & Paarsalu, 1980; Williams, Davids, Burwitz, & Williams, 1993). For example, following Chase and Simon—to examine the domain-specificity effect of expert memory—Allard et al. (1980) administered a recall paradigm to university students. Some of the students were basketball players (i.e., high-skill) and some were not (i.e., low-skill group). Participants viewed slides of structured game information (i.e., offensive basketball play in progress) or unstructured game information (i.e., turnover or rebound). The basketball players recalled the location of

players more accurately than non-players during structured trials only, similar to the work by Chase and Simon (1973). The expert advantage in domain-specific recall of information presented on slides was also demonstrated in women's field hockey (Starkes & Deakin, 1984). In an effort to advance the recall paradigm, Borgeaud and Abernethy (1987) demonstrated higher discrimination between skill groups when dynamic stimuli (e.g., video clips) were employed, in contrast to static stimuli (e.g., slides; cf. Allard & Starkes, 1980; Starkes, 1987).

The use of dynamic stimuli during the recall paradigm has been explored by other researchers. Williams et al. (1993) had experienced and less-experienced soccer players watch 10-second clips of structured or unstructured soccer game-play. Structured trials depicted patterns of play that were common in typical soccer play (i.e., representative of the natural ecology). Unstructured trials contained random patterns of play. After watching clips, participants were required to recall the location of players on-screen. Experienced players made significantly fewer errors in their recall than less-experienced players during structured trials only. In a subsequent test, participants were also assessed on their ability to recognize some of the clips that they previously viewed when presented with both old and new, and also structured and unstructured, video clips. Experienced players demonstrated significantly higher recognition accuracy during only structured trials. Similar to the work in chess (Chase & Simon, 1973) and basketball (Allard et al., 1980), Williams et al. (1993) found that the advantage of experts in recall and recognition



ability only applied to representative trials. Subsequent research has investigated the role of recall and recognition as a predictive measure of skill in sport.

In a subsequent study, Williams and Davids (1995) investigated the recall and recognition ability in high-skill players, low-skill players, and physically disabled spectators who had around 10 years of experience watching soccer—the latter group were included to determine whether recall and recognition skills were developed as a result of experience playing the game or simply experience watching the game. As in Williams et al. (1993), participants viewed structured and unstructured clips of soccer game play and then were required to recall the location of players from the video clip, and subsequently recognize previously-viewed video clips. The high-skill players demonstrated superior *recall* in comparison to both low-skill players and physically-disabled spectators but only on structured trials, whereas they demonstrated superior *recognition* during both structured and unstructured trials.

As part of this study, participants also completed an anticipation task during which they viewed similarly structured clips of soccer game-play and were instructed to anticipate the destinations of passes. High-skill players were significantly better at anticipating the outcome of scenarios than low-skill players. Furthermore, low-skill players were more accurate than the physically-disabled spectators. In a follow-up analysis, Williams and Davids (1995) examined whether these memory skills were predictive of anticipation skill. Recall and recognition performance were highly correlated with anticipation accuracy and skill level, providing convergent validity for the

use of tests of domain-specific memory to assess skill-based performance differences, as measured by anticipation skill and playing experience.

In a similar study, Gorman et al. (2013), had participants complete a traditional recall task (in which they recalled the locations of players after occlusion) and an anticipatory recall task (in which they generated the future position of players after occlusion), in addition to the decision making task (reviewed earlier—see section *Measuring Perceptual-Cognitive Skills in Sport: Differentiating between Skill Groups and Predicting Performance*). Significant skill effects were observed during both traditional and anticipatory recall. Gorman et al. (2013) used a multiple regression model in order to predict decision making skill. Together, accuracy in recall (e.g., traditional, anticipatory) and level of skill accounted for 54.5% of the variance in decision making accuracy, suggesting these constructs are highly related.

In two studies designed to examine the relationship between anticipation and memory, North, Ward, Ericsson, and Williams (2011; see also North, Williams, Hodges, Ward, & Ericsson, 2009) indicated that while there was a significant relationship between anticipation and recognition skill, these skills were supported by separate (albeit similar) processes. Skilled participants demonstrated use of a more complex representation during anticipation than recognition—as evidenced by different eye-movement patterns, and more verbal statements and substantially greater differences in prediction and evaluation statements during anticipation. These data suggest that processes other than recognition alone were responsible for superior anticipation skill and that tests of these skills, rather

than tests of domain-specific memory per se (e.g., recognition and recall, as well as knowledge) may better capture skill-based differences in performance.

**Examining the relative contribution of domain-specific skills and domain-general abilities.** Several studies have not only examined skill-based differences in domain-specific perceptual, cognitive and/or memory skills, such as anticipation, decision making, recall and recognition, but have also examined the contribution of these types of skills relative to other more domain-general visual, perceptual and/or cognitive abilities. In this section, I examine studies that have investigated the relative effectiveness of several domain-specific skills and/or domain-general abilities at differentiating skill groups.

To investigate some of the contributing factors to differences in skill level, Helsen and Starkes (1999) assessed expert and intermediate soccer players' domain-general abilities and domain-specific skills across three experiments. In Experiment 1, the authors measured general perceptual abilities, including measures of processing (e.g., simple reaction time, peripheral reaction time, and visual correction time), optometric properties of vision (e.g., static, dynamic, and mesopic acuity), and peripheral vision (horizontal and vertical peripheral range). In Experiment 2, a variety of domain-specific skills were also recorded, including complex decision speed and accuracy, the number of, duration of, and location of fixations while solving game problems. The results of Experiment 1 demonstrated that expert athletes were not differentiated from less experienced athletes with regards to central and peripheral reaction time, visual correction times, static visual

acuity, visual tracking, or peripheral visual range. The results of Experiment 2 demonstrated that experts were faster and more accurate than less experienced players when deciding what to do next when presented with slides that depicted the first person view of an offensive player with the ball in soccer. In Experiment 3, Helsen and Starkes (1999) assessed expert and intermediate soccer players' ability to respond to 'live-action' scenarios. As the ball approached the point from which the camera was filming, an auditory cue signalled the participant to react by kicking the ball placed in front of them. Decision time and accuracy was recorded in correspondence with participants' actions in response. The results suggested that experts responded faster than novice counterparts. More specifically, experts initiated an action quicker and made contact with the ball sooner. Furthermore, expert players were significantly more accurate in their decisions. Across all experiments, the authors demonstrated that skill-based differences were most prominent in domain-specific skills (e.g., response accuracy, response speed), which explained 84% of the between skill-group variance. Among all of the general abilities assessed, only peripheral visual range contributed significantly, although very weakly (i.e., 3%), to the explanation of between skill-group variance.

In another investigation of skill-based differences in sport, Ward and Williams (2003) examined the development of domain-general visual-perceptual abilities, and domain-specific perceptual-cognitive skills in elite and sub-elite soccer players ranging from 9 to 17 years old. The authors found that visual function, as assessed by dynamic visual acuity, stereoscopic depth sensitivity, and peripheral awareness, did not reliably

differentiate between skill groups at any age. Instead and as in the work of Helsen and Starkes (1999), domain-specific perceptual-cognitive tests reliably differentiated between skill groups. More specifically, tests of players' skill at anticipating the outcome of dynamic situations and assessing the situation were most discriminative of skill level. These results are in line with previous research. First they indicate that while skill groups could be differentiated on most domain-specific measures of perceptual-cognitive skill, two particular measures of domain-specific perceptual-cognitive skill (i.e., anticipation, situation assessment) were more effective in doing so. Second, they indicated that these measures were more discriminating than the domain-general measures used in this research, which were primarily visual in nature.

In general, a number of studies have demonstrated that domain-specific assessments (e.g., response accuracy, anticipation, situation assessment) are better discriminators of skill than domain-general assessments (e.g., visual or physical; cf. Helsen & Starkes, 1999; Ward & Williams, 2003). Contrary to this line of research described above, in a recent meta-analysis of 20 studies, Voss, Kramer, Basak, Prakash, and Roberts (2010) found small, but significant skill-based effects for a number of domain-general and cognitively-based abilities. The authors credit this skill-based effect in 'domain-general' measures to a lack of statistical power in individual studies due to small sample size. When averaged across studies, a moderate effect of skill on domain-general measures reached statistical significance ( $ES = 0.37, p < 0.05$ ). Their analyses indicated that the largest effects were observed in cognitive processing speed ( $ES = 0.67$ ,

$p < 0.05$ ) and attention paradigms ( $ES = 0.53, p < 0.05$ ). These effects were significantly mediated by gender. A moderator analysis revealed that skill-based differences were more profound in males ( $ES = 0.49, p < 0.05$ ) than in females ( $ES = -0.18, p > 0.05$ ). These effects were not significantly moderated by sport type. However, players of interceptive sports ( $ES = 0.71, p < 0.001$ ) accounted for more variance than players of strategic ( $ES = 0.27, p > 0.05$ ) and static ( $ES = 0.10, p > 0.05$ ) sports. In response to domain-general measures explaining between-skill-group variance in sport—particularly cognitive processing speed and attention paradigms—a potential metric of interest to the current research is presented below.

The Berlin Numeracy Test (BNT; Cokely, Galesic, Shulz, Ghazal, Garcia-Retamero, 2012) is a psychometrically sound and quick assessment of risk literacy. Cokely et al. (2012) demonstrate the convergent validity of the BNT through a positive relationship with many general cognitive abilities. For instance, performance on the BNT has been shown to correlate significantly with performance on the Raven's Advanced Matrices test—a psychometric test that assesses cognitive processing speed (Raven, 2000)—and Working Memory Span—a psychometric test for assessing attentional control (Unsworth & Spiller, 2010; see Cokely, et al, 2012). Importantly, and for the current purposes, the BNT can be quickly administered without any loss in predictive validity.

Previously, domain-general perceptual ability has been investigated as a contributor to anticipation skill. However, researchers have either focused on general

measures of perceptual processing (e.g., visual reaction time) or on optometric visual function (e.g., visual acuity). Neither has been shown to differentiate reliably between skill groups (e.g., Helsen & Starkes, 1999; Ward & Williams, 2003). Given the spatial nature of the domain of soccer, a more comprehensive test of perceptual ability that focuses on spatial skills may be a better predictor. The Mental Rotations Test (MRT-A; Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson, 1995) is an AutoCAD-redrawn version of the Vandenberg and Kuse (1978) test of spatial rotation. Wright, Thompson, Ganis, Newcombe, Nora, and Kosslyn (2008) demonstrated that training on the MRT increased the ability to mentally rotate novel stimuli and transferred to another non-practiced spatial task (i.e., a mental paper-folding task), thereby demonstrating the construct validity of the MRT as a measure of spatial ability more generally. For this reason, I plan to use the updated MRT-A as a domain-general measure of spatial ability.

I have now provided a review of the literature that has assessed anticipation and decision making in the lab. Next I reviewed literature that has provided a process-level description of expert performance, particularly that which focuses on option generation. I then provided a review of the literature that has manipulated the effect of time constraint, drawing particular attention to the lack of such research using representative tasks in dynamic domains—despite the increase in temporally-related factors (see Zsombok & Klein, 1997). Lastly, I provided an overview of other constructs that have been used to differentiate between skill groups in sport. This included domain-specific knowledge and memory (e.g., recall, recognition) and some domain-general cognitive measures (e.g.,

processing speed, attention). In order to demonstrate that newly developed tests can explain variance beyond domain-general psychometrics, I have included a description of multiple tests that I will use in addition to the domain-specific assessments of anticipation and decision making. In the next section, I will now provide an overview of the current research and proposed work.

### **Overview of the Current Research**

Multiple models have been proposed that make specific and testable predictions regarding option generation to explain temporally constrained expert decision making (e.g., Johnson & Raab, 2003; Ward et al., 2013). The current research aims to better understand when (i.e., under which conditions) the strategies associated with these models are employed by naturalistic decision makers in soccer game-play situations that are dynamic, time constrained, and uncertain. Multiple experiments are described below. In Experiment 1, I recruited moderately- and less-skilled female soccer athletes to participate in an updated version (i.e., containing newer stimuli) of an established situational option generation task (see Ward, 2002; Ward et al., 2013) that was also adapted to be used during both assessment- and intervention-phases of decision making. The original situational method (Ward, 2002) permits the recording of option-generation behavior (e.g., the number of task-relevant, task-irrelevant, and total situational options generated) and anticipatory performance during the assessment-phase of decision making. In this experiment, this method was extended to also include the generation of



course-of-action options and decision making performance during the intervention-phase of decision making in a similar manner to methods used by Johnson and Raab (2003). The results from this method allow examination of skill effects and the relationship between option generation behavior and performance such that theoretical predictions can be tested. I also examine the effect of manipulating time constraint on option generation behavior and performance in Experiment 1. However, the analyses revealed that skill groups did not differ sufficiently in experience to examine how this factor moderates the interaction between skill and option generation behavior/performance. In Experiment 2, I repeated Experiment 1 using a more conservative time limit but included a low-skill group only as participants.

In Experiment 3, I repeated Experiment 2 using a high-skill group and compared their data to the low skill group in Experiment 2. The results offer considerable insight into the cognitive mechanisms that have been proposed in extant literature to facilitate expert performance (i.e., LTWM, TTF). The effect of time constraint in Experiments 2 and 3 offer some reconciliation between these two models of skilled option generation and provides more insight into which claims of these models are most applicable and when they apply.

In Experiment 4 a shorter, more practically implementable test of perceptual-cognitive skill in soccer was developed—the Online Assessment of Strategic Skill In Soccer (OASSIS). The OASSIS was created using stimuli designed to maximize predictive validity by leveraging the techniques that have been demonstrated to be most

effective in differentiating skill groups in previous research, namely a test of anticipation (as opposed to recall/recognition; see North et al., 2011; 2009) that used temporally occluded stimuli (e.g., Haskins, 1965; Abernethy & Russell, 1987; Abernethy, 1990). The tasks performed during the OASSIS are similar to those presented in the option generation paradigm, albeit using a less time-consuming, multiple-choice format. The purpose of Experiment 4 was threefold. First, performance on the OASSIS provides an independent assessment of the future-oriented component of perceptual-cognitive skill, namely anticipation, with which to compare option generation behavior. Second, known-groups validity is demonstrated by comparing the performance of high- and low-skill groups on OASSIS. Third, predictive validity for OASSIS is demonstrated by examining the relationship between performance on OASSIS and the performance on, and process-level data (i.e., measures of option generation) from, the option generation paradigm within a low-skill group.

In Experiment 5, I repeat Experiment 4 but also demonstrate that the expected relationships exist within an expert group, in addition to a novice group, providing further evidence to support the use of the OASSIS in professional domains. In Experiment 6, I examine the relative contribution of the OASSIS and domain-general cognitive abilities to predicting skill group membership. I demonstrate that the OASSIS is superior to tests of general cognition which were selected based on the findings by Voss et al. (2010). However, I also find that domain-specific knowledge in soccer is superior to differentiating between skill groups than performance on the OASSIS. The implications

of these findings, both for theory and applied utility, are discussed at the end of each experiment and in a more general discussion.

### **Experiment 1<sup>1</sup>**

NCAA Division 3 (i.e., moderate-skill) and recreational-level (i.e., low-skill) female soccer players were recruited to participate in Experiment 1. The purpose of Experiment 1 was to test the claims of TTF and those derived from LTWM (see Ward et al., 2013). To begin, I will review the relevant claims of these models. Proponents of TTF predict that, in specific dynamic, complex, and/or time constrained situations, such as in the types of handball scenarios used in the research by Johnson and Raab (2003), skilled decision makers generate fewer *intervention* options in total than less skilled decision makers (Raab & Johnson, 2007), and the total number of options generated should be negatively related to the quality of performance (i.e., decision quality; Johnson & Raab, 2003). In contrast, based on LTWM, Ward et al. (2013) predicted that, in similar situations (i.e., soccer), skilled decision makers should generate more task-relevant and fewer task-irrelevant *assessment* options than less-skilled decision makers and the number of task-relevant and task-irrelevant options should be positively and negatively related to performance quality (i.e., anticipation accuracy), respectively. In addition to testing these two sets of claims, Experiment 1 provided an initial exploration of the effect of time constraint on performance. Because the models of decision making presented

---

<sup>1</sup> Experiment 1 was also submitted for publication under Belling and Ward (2012; see reference list). Patrick Belling collected and analyzed data, and also wrote drafts of the findings. Dr. Paul Ward assisted with analyses and writing.

above are based on activation by association, it is assumed that, in general, better (i.e., higher quality/semantically meaningful) options should be generated earlier in the sequence of options generated. Therefore, it was logical to assume that the implementation of time constraint would reduce the number of task-irrelevant options generated, but not the number of task-relevant options (i.e., higher quality options, relatively speaking). In accordance with Raab and Johnson (2007), I expected this effect to be more pronounced among players of higher skill. Lastly, I expected that the quality of performance (i.e., anticipation accuracy, decision quality) on the option generation paradigm would differentiate between skill groups since these measures have been shown to reliably discriminate skill levels (e.g., Ward & Williams, 2003; Helsen & Starkses, 1999). Because certain studies on time constraint also suggest that time constraint may affect novices more than experts (cf. Gobet & Simon, 1996a; Calderwood et al., 1988), I expected the moderate-skill group to be less affected by time pressure. Bearing this in mind and recalling that the work done in support of TTF and LTWM was done within different decision perspectives, my specific hypotheses are split into three sections as follows:

### **Hypotheses about Performance during Assessment and Intervention**

*H1.* I expected that that the moderate-skill group will outperform the low-skill group during assessment and intervention trials. Performance during assessment trials was assessed by anticipation accuracy of the actual outcome. To explore performance

during assessment further, I also recorded how often participants generated, accurately rated (as most threatening), and anticipated the criterion most threatening option (rather than actual outcome) for each trial (further details on performance measures are provided in methods below)—because these would be the most semantically meaningful (at least to the moderate-skill group) but not necessarily the most likely option. Performance during intervention trials was assessed by how often participants selected the criterion best option for each trial. To explore performance during intervention further, I also recorded how often participants generated and accurately rated the criterion best option. I expected to observe a main effect of time constraint (within participants: present/absent), a main effect of skill (between participants: moderate/low) and a Skill x Time Constraint interaction on performance measures. While both groups were expected to be affected negatively by time constraint and the moderate-skill group was expected to be more accurate than the low-skill group, I expected the moderate-skill group to be less affected by time constraint than the low-skill group.

### **Hypotheses about Process during Assessment**

*H2.* During assessment trials, I expected that the moderate-skill group would generate more task-relevant and fewer task-irrelevant options than the low-skill group (see Ward et al., 2013). I expected this effect would be stronger when under time constraint. Accordingly, I expect to observe a 3-way interaction of Skill (between participants: moderate/low) x Information Type (within participants: task-relevant/task-

irrelevant) x Time Constraint (within participants: present/absent) on option generation behavior measures. More specifically, the nature of the Skill x Information Type interaction was expected to be substantially different at each level of time constraint. Moderate-skilled participants were expected to utilize an adaptive (rather than maladaptive) strategy and generate fewer task-irrelevant options during time constrained trials. In comparison, it was less clear that less-skilled participants would employ a similar adaptive strategy—at least to the extent expected of the moderate-skilled group.

*H3.* During assessment trials, I expected positive and negative relationships between performance quality and the number of task-relevant and task-irrelevant options generated, respectively (see Ward et al., 2013). How time constraint would affect these relationships was largely exploratory. Since time constraint was expected to reduce the number of task-irrelevant options generated primarily, I speculated that the presence of time constraint would result in a stronger negative correlation between the number of task-irrelevant options generated and performance. In other words, generating more task-irrelevant options would be more detrimental to anticipation accuracy in time constrained conditions than when not under time constraint. I also speculated that the positive correlation between the number of task-relevant options generated and performance would remain largely unaffected by time pressure, especially for the moderate-skill group.

### **Hypotheses about Process during Intervention**

**H4.** During intervention trials, I expected that the moderate-skill group would generate fewer total options than the low-skill group (Raab & Johnson, 2007). These effects were expected to be stronger when under time constraint. To explore the effects observed by Ward et al. (2013)—and the expectation that time constraint would largely affect the generation of task-irrelevant options (especially for more skilled players)—I also examined whether these effects would be moderated by information type. I speculated that I would observe a 3-way interaction of Skill (between participants: moderate/low) x Information Type (within participants: task-relevant/task-irrelevant) x Time Constraint (within participants: present/absent), wherein the moderate-skill group would generate more task-relevant and fewer task-irrelevant options than the low-skill group under no time constraint (see Ward et al, 2013). Additionally, I speculated that this difference would be greater under time constraint because the moderate-skill group would generate even fewer task-irrelevant options. In contrast, skill main or interaction effects in the absence of an information type main or interaction effect would lend support for TTF.

**H5.** During intervention trials, I expected to observe a negative relationship between the total number of options generated and performance (see Johnson & Raab, 2003). I expected this relationship would be stronger during time constrained trials—given the simple heuristic’s focus on generating an immediately satisficing option. To further explore the information type effects observed by Ward et al. (2013)—and the expectation that time constraint would largely affect the generation of task-irrelevant

options—separate correlations were conducted for each condition. I speculated that time constraint would result in a stronger negative correlation between the number of task-irrelevant options generated and performance compared to when time to generate options was not constrained. I also speculated that the positive correlation between the number of task-relevant options generated and performance would remain largely unaffected by time pressure, especially for the moderate-skill players. In other words, the generation of additional irrelevant options would be more negatively correlated with performance during time constrained trials than performance during non-time constrained trials.

## **Methods**

### **Participants**

The moderate-skill group was composed of 23 female NCAA Division II soccer players with a mean age of 18.70 ( $SD = 0.70$ ) and 10.41 ( $SD = 2.00$ ) years of experience playing competitive soccer (i.e., organized tournament match-play under the supervision of a coach). The low-skill group was composed of 13 female recreational-level players with a mean age of 19.62 ( $SD = 1.66$ ) and 10.00 ( $SD = 5.03$ ) years of experience. Participation in the research was completely voluntary.

### **Materials**

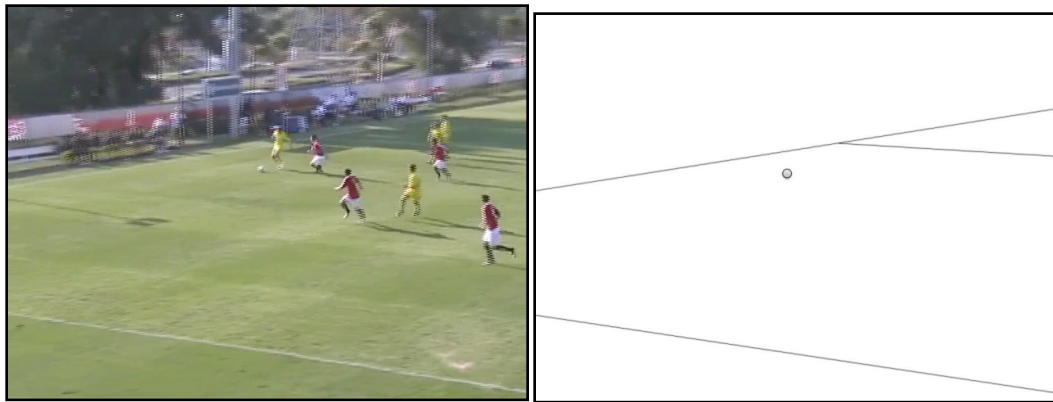
Thirty video simulations were created using video footage provided by the U.S. Soccer Federation. The video footage was of live, national-level, inter-academy games depicting full-team play (i.e., 11-player vs. 11-player). When discriminating between



skill groups, Ward and Williams (2003) found nearly double the effect size when using full-team scenarios ( $ES = 0.95$ ) than when using 1-on-1 scenarios ( $ES = 0.50$ ). The footage used in the present study was filmed from an elevated perspective, above and behind the goal, displaying the entire length of the field, but allowing for a pseudo first-person perspective. Previous research has demonstrated the effectiveness of this viewpoint at differentiating between skill groups (see Mann, Farrow, Shuttleworth, Hopwood, & MacMahon, 2009). All simulations were displayed using a high-definition projector. The image was displayed on a reflective-painted wall and was 249 centimeters wide by 158 centimeters in height.

In assessment trials ( $n = 12$ ), a team with possession of the ball was attacking from the far side of the field (i.e., the top of the screen) towards the near end of the field (i.e., the bottom of the screen). This view facilitated a defensive perspective. In intervention trials ( $n = 12$ ), a team with possession of the ball was attacking towards the far side of the field (i.e., the top of the screen). This view facilitated an offensive perspective. Of the remaining six video simulations, three were used as training trials for the assessment task and three were used as training trials for the intervention task. All video simulations featured a few seconds (e.g., 5-10s) of dynamic build-up play to provide context to the impending task. The simulations ended unexpectedly and were occluded (see Figure 1; Farrow et al., 2005; Ward et al., 2013) immediately prior to a critical decision moment when the player with the ball was about to take an action (e.g., shoot, pass, dribble with the ball). The occlusion image displayed a blank white screen

with only black lines to represent the field lines and location of the ball. All perceptual information involving the players on the field was removed. Previous research has shown that leaving additional context on screen (i.e., a jpeg of the final frame of action) as participants perform the experimental tasks inflates the number of options generated and decreases the difficulty of the tasks (see Ward, et al., 2013).

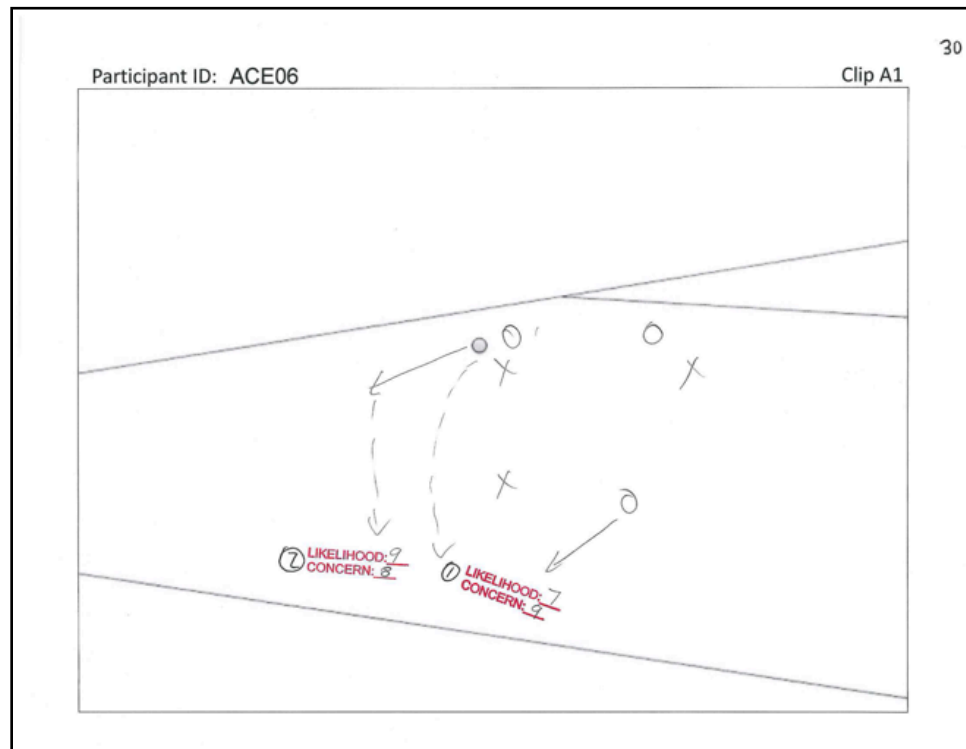


**Figure 1:** The final frame of a video simulation (left) and the matching occlusion image (right) during the option generation paradigm.

Response sheets, for both assessment and intervention trials, were created on standard letter-sized (8- by 11-inch) white paper to match the perspective of the occlusion image on-screen. Separate custom stamps were created for assessment and intervention trials. Stamps created for use during assessment trials displayed “LIKELIHOOD:\_\_\_\_\_” and “CONCERN:\_\_\_\_\_” in red ink on the response sheet (see Figure 2). Stamps created for use during intervention trials displayed “LIKELIHOOD:\_\_\_\_\_” and “QUALITY:\_\_\_\_\_” in blue ink on the response sheet (see procedure of further

explanation). A stopwatch was used to implement time constraint on half of all scored trials.

Additionally, as a manipulation check for the effects of time constraint, the Mental Readiness Form (MRF-L; see Cox, Russell, & Robb, 1999 for full review) was used to assess state-anxiety while completing the trials from each perspective. The MRF-L was constructed to provide an accurate, but brief assessment of state anxiety during sport competition and is considered a valid psychometric test for this purpose (see Cox, Russell, & Robb, 1999). The Rating Scale Mental Effort (RSME; see Zijlstra, 1993 for full review) was also used to provide a measure of mental effort for the same purpose. Lastly, a questionnaire was created to assess the relevant soccer playing and watching experience of participants (see Figure 7 in appendix).



**Figure 2:** A matching response sheet (to occlusion image in Figure 1) with written and stamped options and ratings of likelihood and concern during assessment on the option generation paradigm.

## **Procedure**

For assessment trials, participants were instructed to watch the video simulations and envision themselves as members of the defensive team. They were told the opposing team would be attacking the goal at the bottom of the screen, with which it was their job to defend. Next, they were told each video simulation would end at a moment in which a player on the opposing team would have the ball and would be about to make a decision about what to do next. Participants were instructed to write down all of the options that they would be concerned about the opponent with the ball doing next, as if they were playing defense on the field. They were specifically instructed not to write down all of the options available to the player that existed in the environment, which could be an extremely exhaustive list. Instead, they were informed only to write down those options that they were actually thinking about when the video clip was occluded. Participants wrote options onto the paper using a simple notation scheme (see Figure 2). On assessment trials, a letter ‘X’ marked on paper represented a defensive player—and a teammate. A letter ‘O’ represented an offensive player—and member of the opposing team. Solid arrows were used to indicate player movement with or without the ball (e.g., the player with the ball dribbling in a direction, a player off-the-ball running into a space

to receive it). Once they had written down all of the options participants were instructed to mark each option with the custom stamp made for assessment trials. After stamping, participants rated the likelihood of each option to be chosen by the player with the ball on the opposing team and the level of concern (i.e., threat) they felt about each option as a member of the defensive team. Participants rated the likelihood of each option using a Likert-type scale that ranged from 0 (not at all likely) to 10 (very likely). Participants also rated their concern regarding each option using a Likert-type scale that ranged from 0 (not at all concerned) to 10 (very concerned).

For intervention trials, participants were instructed to watch the video simulations and envision themselves on the attacking team, and more specifically, as the player with the ball at the moment of occlusion on-screen. They were informed each clip would end with their team attacking the goal at the far end of the pitch, which was at the top of the screen. They were instructed to generate the intervention options that they considered choosing as that player. Using the same notation described above—with the exception that the letter ‘X’ represented the opposing players (i.e., defenders) and the letter ‘O’ represented teammates (i.e., attackers) during intervention trials—participants wrote their options down on the response sheet. After all options were generated, each option was stamped and rated using the custom stamp for intervention trials. After stamping, participants rated the likelihood with which they would pursue that option, and the quality of that option (i.e., how good they felt that option was for them, given the current situation). Again, participants rated the likelihood using a Likert-type scale from 0 (not at

all likely) to 10 (very likely). The Likert-type scale for the quality of each option ranged from 0 (not at all good) to 10 (very good).

After consent was obtained, participants received the instructions described above and completed two training trials for either assessment or intervention under the guidance of an experimenter. During these trials, the experimenter would read options back to the participant in order to ensure the options conceived at the time of occlusion matched what was drawn on paper. A third training trial was completed under time constraint to familiarize participants to the time pressure implemented during test trials. During half of the 24 test trials (i.e., 6 assessment and 6 intervention trials), participants were given only 15 seconds (post occlusion) to generate and mark options down on paper. Participants were instructed that they could provide accurate ratings of likelihood and concern/quality after the time limit. They were also permitted to ‘clean-up’ options that were illegible on the response sheet. However, participants were not permitted to add (i.e., generate more) options or change the functional meaning of any of the options that they had written down. In the remaining half of the trials, participants generated options without time constraint.

All four conditions (i.e., time constrained assessment, non-time constrained assessment, time constrained intervention, and non-time constrained intervention trials) were experimentally counterbalanced to avoid any effects of task order. As a manipulation check to assess the effect of time pressure in each condition, participants were instructed to complete the MRF-L form and RSME form immediately after and in

reference to the last six trials only. Once during the session, after exactly 12 trials (i.e., during the task change in decision perspective), the experimenter administered a soccer experience questionnaire.

## **Data Analysis**

**Coding of options on response sheet.** Two subject-matter experts (SMEs) were used in the coding of options and determining of options' task-relevance. SME1 had 18 years of playing experience, including a year as captain of a NCAA Division III collegiate level. SME2 had 17 years of playing competitive soccer, including multiple years as a collegiate club-level captain. Both SMEs had the advantage of watching video clips multiple times and were not limited to only the occluded version. SMEs also had the advantage of watching a non-occluded version of the stimuli and the subsequent sequence of play following the critical decision point.

Participants' responses were categorized based on the action, direction, and location of players and ball movement. Options on the response sheet were also coded functionally (i.e., a pass between two defenders would be deemed a through-pass even if the location of the players was not perfectly accurate). SME1 coded all options generated by participants. A portion (approx. 17%) of the responses was coded independently by SME2. During assessment trials, inter-observer agreement in terms of functional coding was 85 percent (Cohen's Kappa = 0.82). During intervention trials, inter-observer

agreement was 82 percent (Cohen's Kappa = 0.66). The SMEs discussed options where there was disagreement until agreement was reached.

In order to categorize generated options into 'task-relevant' and 'task-irrelevant' (see Ward et al., 2013), both SMEs generated what they thought to be the task-relevant options for each assessment and intervention trial. Agreement between SMEs in terms of task-relevance was 97 percent (Cohen's Kappa = 0.96) across all trials. Those on which they disagreed were discussed until agreement was reached. Whenever a participant generated an option that matched one of the task-relevant options generated by the SMEs, it was deemed to be task-relevant. Any other options generated were deemed to be task-irrelevant.

**Measures.** To assess performance during assessment, anticipation accuracy of the actual outcome was recorded. Anticipation of outcome was determined by the correspondence between the participant's option with the highest likelihood rating and the actual outcome. To assess performance during intervention, the selection of the criterion best option were recorded. The criterion best option was defined as the option deemed by the SMEs to be the highest quality course of action that could be taken at that point in the game. The frequency with which this option was selected by each participant as their best one was measured.

To explore performance further, additional dependent variables were measured—both for assessment and intervention (reported in Tables 13 and 14 in appendix). During assessment, I recorded the generation, accurate rating, and anticipation of the criterion



most threatening option were recorded as additional measures of performance. The criterion most threatening option was defined as the option rated of most concern to the defense by the SMEs. The frequency with which this was generated by the participant was measured (irrespective of whether that option was rated by the participant as most threatening or anticipated to be the actual outcome). The frequency with which this was generated *and* accurately rated as the most threatening, and the frequency with which this was generated *and* anticipated to be the actual outcome were also measured. During intervention, I recorded the generation, and accurate rating of the criterion best option. The frequency with which the criterion best option was generated by the participant was measured (irrespective of whether that option was rated by the participant as their best option, or selected as the next course of action). The frequency with which this was generated *and* accurately rated as the criterion best option was also measured. To assess option generation behavior during assessment, the number of total options, number of task-relevant options, and number of task-irrelevant options generated for each trial were recorded. To assess option generation behavior during intervention, the number of total options, number of task-relevant options, and number of task-irrelevant options generated for each trial were recorded.

**Analysis.** Two Skill x Time Constraint factorial ANOVAs were conducted to examine the hypotheses that the moderate-skill group would (a) anticipate more accurately and (b) select the criterion best response option more frequently than the low-skill group, particularly when under time constraint, during both assessment and

intervention, respectively (*H1*). Skill (moderate/low) was used as a between-participants variable. Time constraint (present/absent) was a within-participants independent variable. Similar analyses were performed on the additional dependent variables to explore differences in performance further (see previous section). Skill effect sizes were calculated using a partial eta squared explanation of variance ( $\eta p^2$ ).

A 3-way (Skill x Time Constraint x Information Type) ANOVA was used to test the hypothesis that the moderate-skill group would generate more task-relevant and fewer task-irrelevant options than the low-skill group during assessment trials (see Ward et al., 2013), and the reduction in task-irrelevant options would be particularly pronounced for the moderate-skill group under time constraint (*H2*). Information type (task-relevant/task irrelevant options) was included as an additional within-participant factor. As a follow up, I used simple interaction effects (see Howell, 2002) to examine the relationship between information type and skill level on option generation within each time constraint condition.

Pearson's Product Moment Coefficient ( $r$ ) was used to examine the hypotheses that, during assessment trials, the number of task-relevant and task- irrelevant options generated would be positively and negatively correlated with anticipation accuracy, respectively, and these would be moderated by time constraint (*H3*). Recall that a stronger negative relationship between task-irrelevant information and anticipation accuracy was expected during time constrained trials than during non-time constrained trials, and that the relationship between the number of task-relevant options and

anticipation would remain unaffected by time constraint, particularly within the moderate-skill group. Separate correlations were initially conducted between each condition (i.e., task-irrelevant/time constrained; task-irrelevant/non-time constrained; task-relevant/time constrained; task-relevant/non-time constrained) and anticipation accuracy. Since these analyses were under powered, I also examined this relationship across both time constraint conditions and skill groups combined—to gain an idea of which cognitive mechanism was being employed more generally. In order to explore the claims of TTF, I also observed the relationship between the total number of options and performance within and across all time conditions. If the total number of options was more predictive of performance than the number of task-relevant/task-irrelevant options, support for TTF would be provided.

A 3-way (Skill x Time Constraint x Information Type) ANOVA was conducted to test the hypothesis that, during intervention trials, the moderate-skill group would generate fewer total options (Raab & Johnson, 2007) than the low-skill group, particularly when under time constraint. This analysis also allowed the proposal to be tested that while generating fewer in total, the moderate-skill group would generate more task-relevant and fewer task-irrelevant options (see Ward et al., 2013) and would reduce the number of task-irrelevant options when under time constraint. As a follow up, I used simple interaction effects (see Howell, 2002) to examine the relationship between information type and skill level on option generation within each time constraint condition. In contrast, a skill effect or interaction in the absence of an information type

main effect or interaction would lend support for TTF because it would indicate that the moderate-group only generated fewer options in total than the low-skill group.

During intervention trials, I hypothesized that the total number of options generated would be negatively correlated with performance, and that this effect would be exaggerated by the presence of time constraint (*H5*). I used Pearson's *r* to observe the relationship between the total number of options and selection of the criterion best option within each time constraint condition. Also, to test the claims of Ward et al. (2013), I used Pearson's *r* to observe the relationships between the number of task-relevant (and task-irrelevant) options generated and selection of the criterion best option. To explore the prediction that time constraint would affect the generation of mostly task-irrelevant information, I observed the relationship between the number of task-irrelevant options generated and performance within each time condition. A stronger negative relationship between task-irrelevant options and performance during time constrained trials than non-time constrained trials would indicate support for the hypothesis. I also hypothesized that the relationship between task-relevant options and performance would remain relatively unaffected by time constraint, particularly within the moderate-skill group. To explore this, I observed the relationships between the number of task-relevant options and performance within each time condition and skill group. Due to low power, I also observed the correlation between the number of task-relevant/task-irrelevant/total options on average across both time conditions and performance. Across all intervention trials, stronger positive/negative relationships between performance and task-relevant/task-

irrelevant options than between performance and total options would be taken to indicate that LTWM-type mechanisms were being employed, as opposed to the TTF heuristic.

As a manipulation check (i.e., to observe the effect of time constraint on self-assessed anxiety and effort ratings), scores on the MRF-L and RSME were analyzed using a factorial ANOVA, with time constraint as the within participants variable and skill level as the between participants variable. Main effects of time constraint would indicate that time pressure was effective in pressuring participants.

Cohen's  $d$  was used as an estimator of between-participants effect size on the number of options generated. Within-participants effect size ( $ES$ ) on the manipulation check were calculated by expressing the mean difference in standard units, using the baseline, non-time constrained condition SD as the denominator (see Hedges & Olkin, 1985; Rosenthal & DiMatteo, 2001). Partial eta squared was used as an estimator of effect size on all performance and 3-way interaction effects. For simple interaction effects within each time setting, Cohen's  $d$  was calculated for both types of information (task-relevant and -irrelevant) generated by both skill groups.

## **Results**

### **Manipulation Check**

During assessment trials, participants reported significantly higher levels of anxiety when under time constraint than when under no time constraint,  $F(1, 34) = 5.16$ ,  $p = 0.03$ ,  $ES = 0.30$ . However, participants did not report significantly higher levels of

effort when under time constraint than when under no time constraint,  $F(1, 34) = 3.23$ ,  $p = 0.08$ ,  $ES = 0.26$ . For a full review of the effect of time constraint on anxiety and mental effort during assessment trials, see Table 1.

**Table 1.** Anxiety and mental effort ratings during time constrained and non-time constrained assessment trials of Experiment 1.

Manipulation check	Anxiety rating		Mental effort rating	
	Time condition	Time constraint      No time constraint	Time constraint	No time constraint
Mean (SD)		14.72 (6.02)      12.75 (6.57)	62.22 (23.43)	55.69 (25.50)
Effect of time constraint		$F(1, 34) = 5.16$ $p = 0.03$ , $ES = 0.30$	$F(1, 34) = 3.23$ $p = 0.08$ , $ES = 0.26$	

During intervention trials, participants reported significantly higher levels of anxiety when under time constraint than when under no time constraint,  $F(1, 34) = 14.11$ ,  $p < 0.01$ ,  $ES = 0.40$ . Furthermore, participants rated significantly higher levels of effort when under time constraint than when under no time constraint,  $F(1, 34) = 5.38$ ,  $p = 0.03$ ,  $ES = 0.16$ . For a full review of the effect of time constraint on anxiety and mental effort during intervention trials, see Table 2.

**Table 2.** Anxiety and mental effort ratings during time constrained and non-time constrained intervention trials of Experiment 1

Manipulation check	Anxiety rating		Mental effort rating	
	Time condition	Time constraint      No time constraint	Time constraint	No time constraint
Mean (SD)		15.36 (5.17)      13.25 (5.33)	62.78 (22.69)	58.61 (26.50)
Effect of time		$F(1, 34) = 14.11$	$F(1, 34) = 5.38$	

constraint	$p < 0.01, ES = 0.40$	$p = 0.03, ES = 0.16$
------------	-----------------------	-----------------------

## Performance

**H1.** During assessment, I hypothesized that the moderate-skill group would outperform the low-skill group, particularly during time constrained trials. Accordingly, I expected to observe a main effect of skill and a 2-way interaction of Skill x Time Constraint on performance. A main effect of skill was observed  $F(1, 34) = 17.97, p < 0.01$  (see Table 3). The moderate-skill group anticipated the outcome of scenarios more accurately than the low-skill group during both time- and non-time constrained trials. However, a Skill x Time Constraint interaction was not observed (for a full review of performance during assessment, see Table 13 in appendix).

During intervention, (**H1**) I hypothesized that the moderate-skill group would outperform the low-skill group, particularly during time constrained trials. I expected to observe a main effect of skill and a 2-way Skill x Time Constraint interaction effect on performance. Counter to this hypothesis, a main effect of skill was not observed (see Table 3). The moderate-skill group did not select the criterion best option of scenarios more often than the low-skill group  $F(1, 34) = 1.44, p = 0.24$ . Also, a Skill x Time Constraint interaction was not observed (for a full review of performance during intervention, see Table 14 in appendix).

**Table 3.** Mean (SD) frequency scores for anticipation accuracy during six time constrained and six non-time constrained assessment trials and selection of the criterion

best option during six time constrained and six non-time constrained intervention trials, main effects of skill and time constraint, and interaction effects of skill by time constraint on performance during assessment and intervention trials of Experiment 1.

	Anticipation accuracy		Selection of criterion best option	
	Time constraint	No time constraint	Time constraint	No time constraint
Low-skill	0.85 (0.55)	0.85 (0.69)	1.31 (1.11)	1.46 (0.97)
Moderate-skill	1.39 (0.99)	1.39 (0.78)	1.87 (1.28)	1.56 (1.20)
Skill effect	$F(1, 34) = 17.97$ $p < 0.01, \eta^2 = 0.35$		$F(1, 34) = 1.44$ $p = 0.24, \eta^2 = 0.05$	
Time constraint effect	$F(1, 34) = 0.00$ $p = 1.00, \eta^2 = 0.00$		$F(1, 34) = 0.06$ $p = 0.81, \eta^2 < 0.01$	
Skill x Time interaction effect	$F(1, 34) = 0.00$ $p = 1.00, \eta^2 = 0.00$		$F(1, 34) = 0.53$ $p = 0.47, \eta^2 = 0.02$	

## Option Generation during Assessment

**H2.** During assessment, I hypothesized that the moderate-skill group would generate more task-relevant and fewer task-irrelevant options, particularly during time constrained trials. Accordingly, I expected to observe a 3-way interaction of Skill x Information Type x Time Constraint—and a different type of 2-way interaction at each level of time constraint. The 3-way interaction was not significant  $F(1, 34) = 1.12, p = 0.30, \eta^2 = 0.03$ . During time constrained trials only, a Skill x Information Type simple interaction effect was observed,  $F(1, 34) = 4.70, p = 0.04$  (see Table 4 for effect sizes within each time setting). The moderate-skill group generated more task-relevant and fewer task-irrelevant options than the low-skill group. During non-time constrained trials, a Skill x Information Type simple interaction effect was not observed,  $F(1, 34) = 0.45, p = 0.51$  (see Table 4 for effect sizes within each time setting). In general during



assessment trials, the moderate-skill group reacted to time constraint by reducing the generation of task-irrelevant information only, whereas the low-skill group reduced the generation of both task-relevant and task-irrelevant information.

**Table 4.** Means (SD) number of options generated per trial and interaction effects of skill by information type on number of options generated during assessment trials under each time constraint condition of Experiment 1.

	Information type	Skill group	Trials with time constraint	Trials with no time constraint
Number of options generated	Task-relevant	Low	0.78 (0.46)	0.91 (0.41)
		Moderate	1.05 (0.39)	1.10 (0.45)
		Cohen's <i>d</i>	<i>d</i> = 0.63	<i>d</i> = 0.44
	Task-irrelevant	Low	1.19 (0.51)	1.33 (0.57)
		Moderate	0.88 (0.51)	1.32 (0.68)
		Cohen's <i>d</i>	<i>d</i> = -0.61	<i>d</i> = -0.02
Skill x Information Type simple interaction effect			<i>F</i> (1, 34) = 4.70 <i>p</i> = 0.04	<i>F</i> (1, 34) = 0.45 <i>p</i> = 0.51

**H3.** Recall that I hypothesized, during assessment, (a) the number of task-relevant and task-irrelevant options generated to be positively and negatively related to anticipation accuracy, respectively (Ward et al., 2013) (but the total number of options generated would not be correlated with performance, cf. Johnson & Raab, 2003) and (b) a stronger negative relationship between the number of task-irrelevant options and performance during time constrained trials than during non-time constrained trials. Both the number of task-relevant options and task-irrelevant options were not significantly correlated with anticipation accuracy—in either direction—during either time constraint condition. In order to observe which model (e.g., LTWM, TTF) tended to receive more

support, correlations were observed across time conditions and skill groups. The following data are therefore presented very tentatively.

When averaged across both time conditions, the number of task-relevant options generated was significantly and positively correlated with anticipation accuracy ( $r = 0.35$ ,  $p = 0.04$ ). The relationship between the number of task-irrelevant options generated and anticipation accuracy approached significance in the negative direction ( $r = -0.31$ ,  $p = 0.06$ ), lending some tentative support for the LTWM proposal outlined by Ward et al. (2013). Support for TTF during assessment was not observed. The total number of options was not correlated with anticipation accuracy within any time condition. Even when averaged across time conditions, the total number of options generated was not significantly negatively related to anticipation accuracy ( $r = -0.08$ ,  $p = 0.64$ ).

Next, because no significant correlations existed between the number of task-relevant/task-irrelevant options and performance *within* time conditions (see paragraph above), no support was offered for the hypothesis that a stronger negative correlation would exist between task-irrelevant options generated and performance during time constrained trials than during non-time constrained trials ( $H3$ ). I also hypothesized that the number of task-relevant options would be related to performance regardless of time constraint condition, particularly within the moderate-skill group. Within the moderate-skill group, the number of task-relevant options was not related to anticipation accuracy during time constrained trials ( $r = -0.02$ ,  $p = 0.95$ ) or during non-time constrained trials ( $r = -0.01$ ,  $p = 0.96$ ). Also, within the low-skill group, the number of task-relevant options

was not related to anticipation accuracy during time constrained ( $r = 0.24, p = 0.44$ ) or non-time constrained trials ( $r = -0.15, p = 0.62$ ). Therefore, no support was observed for this portion of the hypothesis (*H3*).

### **Option Generation during Intervention**

*H4.* Recall that during intervention trials, I hypothesized that the moderate-skill (compared to low-skill) group would generate fewer total options, particularly under time constraint. However, I also wanted to explore the claims of Ward et al. (2013) who predicted that higher skilled individuals would generate more task-relevant and fewer task-irrelevant options. While Ward et al. (2013) did not make predictions regarding time constraint, I predicted that this would be particularly pronounced during time constrained trials. Accordingly, I explored whether there was a significant Skill x Information Type x Time Constraint interaction—wherein a different type of 2-way interaction was expected at each level of time constraint. A skill effect or interaction in the absence of an information type main effect or interaction would lend support for TTF.

The 3-way interaction was not significant  $F(1, 34) = 0.45, p = 0.51, \eta p^2 = 0.01$ . However, a significant Skill x Information Type simple interaction effect was observed during time constrained trials,  $F(1, 34) = 5.06, p = 0.03$  (see Table 5 for effect sizes within each time setting). The moderate-skill group generated more task-relevant and fewer task-irrelevant options during time constrained trials than the low-skill group. During non-time constrained trials, a Skill x Information Type simple interaction effect

was not observed,  $F(1, 34) = 1.69, p = 0.20$  (see Table 5 for effect sizes within each time setting). In general, these simple interaction effects suggest that the moderate-skill group reacted to time constraint during intervention trials by reducing the generation of task-irrelevant information only, whereas the low-skill group reduced the generation of both task-relevant and task-irrelevant information.

**Table 5.** Means (SD) number and type of options generated per trial and simple interaction effects of skill by information type on number of options generated during intervention trials under each time constraint condition of Experiment 1.

	Information type	Skill group	Trials with time constraint	Trials with no time constraint
Number of options generated	Task-relevant	Low	0.77 (0.28)	0.88 (0.31)
		Moderate	0.96 (0.31)	1.06 (0.33)
		Cohen's $d$	$d = 0.64$	$d = 0.56$
	Task-irrelevant	Low	1.28 (0.38)	1.37 (0.52)
		Moderate	1.07 (0.51)	1.32 (0.43)
		Cohen's $d$	$d = -0.47$	$d = -0.10$
Skill x Information Type simple interaction effect			$F(1, 34) = 5.06$ $p = 0.03$	$F(1, 34) = 1.69$ $p = 0.20$

**H5.** Recall that I hypothesized that the total number of options would be negatively related with performance (i.e., selection of the criterion best option) during intervention trials. I expected to observe a stronger correlation during time constrained than non-time constrained trials. The total number of options generated was not significantly correlated with performance during time constrained ( $r = -0.16, p = 0.37$ ) or non-time constrained ( $r = -0.15, p = 0.39$ ) trials. As in assessment, I averaged the relationships across both time conditions in order to increase power. When averaged

across both time constraint conditions, counter to the TTF-based hypothesis the total number of options generated was still not significantly negatively related to selecting the criterion best option ( $r = -0.09, p = 0.62$ ). Therefore, support was not found for this part of the hypothesis (*H5*) or for TTF.

Additionally, and during intervention, I explored the claims of Ward et al. (2013) regarding the relationships between performance and task-relevant/task-irrelevant options, and whether these would be more pronounced under time constraint. I hypothesized that the number of task-irrelevant options and performance would be more negatively correlated during time constrained trials than during non-time constrained trials. Within time constrained trials, the number of task-irrelevant options generated was negatively related to performance ( $r = -0.37, p = 0.03$ ). Within non-time constrained trials, the relationship between the number of task-irrelevant options and performance approached significance ( $r = -0.31, p = 0.07$ ). This indicated support for this portion of the hypothesis (*H5*).

During intervention trials, I also hypothesized that the relationship between task-relevant options and performance would remain relatively unaffected by time constraint, particularly within the moderate-skill group. Within the moderate-skill group, the number of task-relevant options was not significantly related to the selection of the best option during time constrained ( $r = 0.20, p = 0.37$ ) or during non-time constrained ( $r = 0.15, p = 0.50$ ) trials. Within the low-skill group, the number of task-relevant options was not related to selection of the best option during time constrained ( $r = 0.39, p = 0.20$ ) or non-

time constrained ( $r = 0.29, p = 0.34$ ) trials. Therefore, no support was observed for this portion of the hypothesis ( $H5$ ).

When averaged across both time constraint conditions (as in assessment trials, see  $H3$ ), the relationship between selecting the criterion best option and the number of task-relevant ( $r = 0.27, p = 0.11$ ) and task-irrelevant ( $r = -0.32, p = 0.06$ ) options generated approached significance in the directions hypothesized by Ward et al. (2013).

## **Discussion**

I expected that the moderate-skill group would outperform the low-skill group ( $H1$ ). During assessment, this hypothesis was supported. The moderate-skill group anticipated the outcome of trials more accurately than the low-skill group. Two additional performance measures also reached or approached significance (see Table 13 in appendix). However, similar effects were not observed during intervention. The moderate-skill group did not select the criterion best option significantly more often than the low-skill group. The additional exploratory data (see Table 14 in appendix) did demonstrate that they generated the criterion best option significantly more often than the low-skill group. The skill effect for accurately rating the criterion best option also approached significance (see Table 14 in appendix). Since the separation between groups in terms of skill level was not extensive, it is possible that training at this level focuses on the anticipation of opponents' actions more than the ability to quickly decide upon one's own actions—or that assessment skills develop before those deciding how to intervene

effectively. Future research should design separate training programs for each of these phases of decision making and test their effectiveness in relation to current training techniques used on the field.

With regard to performance, I also hypothesized that presence of time constraint would affect the moderate-skill group less than the low-skill group (see Gobet & Simon, 1996a; Calderwood et al., 1988). Instead, a Time Constraint x Skill effect was not observed on performance during assessment or intervention trials (see Chabris & Hearst, 2003; Holding, 1989; Dunne, 1995). The absence of a time constraint main effect suggests that this manipulation may not have been effective in negatively affecting, at least, low-skill performance (but see option generation data). It is possible that performance on a representative task in a dynamic domain, such as soccer, may rely on different mechanisms than performance on representative tasks in more traditional domains, such as chess—where time has been shown to negatively affect performance, especially in low-skilled individuals. While novice chess players may rely primarily on search-based mechanisms, the dynamic nature of soccer may require *all* players to rely on cognitive mechanisms that facilitate quick and satisficing option generation, such as those investigated in this research (e.g., LTWM, TTF), thereby reducing the effect of time constraint on performance.

However, there are two limitations in the research that may also have contributed this finding. First, the skill groups were not well differentiated in Experiment 1. The moderate-skill group was composed of entirely NCAA Division 3 athletes and the low-

skill group of entirely recreational-level players, including ex-high school varsity and collegiate club level players. The moderate-skill group only averaged 0.41 more years of experience than the low-skill group. Future research should investigate these issues with more well-defined and differentiated skill groups to address this issue. Second, the effect of time constraint may not have been sufficient to affect participants. Participants were given 15 seconds to respond during time constrained trials. However, while a moderate effect size was observed in anxiety (i.e., participants were moderately more anxious under time constraint) participants often finished within 15 seconds during non-time constrained trials possibly suggesting that there was no real effect on performance. Future research should use a more conservative time limit during time constrained trials (e.g., <15s).

During assessment trials I hypothesized that the moderate-skill group would generate more task-relevant and fewer task-irrelevant options than the low-skill group, particularly during time constraint (*H2*). While the hypothesized 3-way interaction was not observed, the simple interaction effects indicated that in time constrained trials only, the moderate-skill group generated significantly more task-relevant and less-task relevant options than the low-skill group. This data may suggest that while moderate-skill participants reduced the generation of task-irrelevant information only when under time pressure, low-skill participants simply reduced the amount of all information generated, irrespective of its relevance to the situation. The reduction in the generation of unnecessary information by the moderate-skill group (as opposed to the low-skill group)



is in line with the claims of Raab and Johnson (2007) and has been documented elsewhere as a potential mechanism of skill acquisition (see Haider & Frensch, 1999). In other words, the results suggest that as more skill is acquired, better options are generated quickly and unnecessary options are reduced (see Raab & Johnson, 2007). While this finding is in line with the TTF heuristic, other interaction effects provide further insight into the potential mechanism supporting skilled performance.

In general, the Skill x Information Type interaction effect indicates that the categorization of options into task-relevant and task-irrelevant may be more useful for differentiating between skill groups (see Ward et al., 2013). In other words, the number of task-relevant and number of task-irrelevant options differentiated between skill during time constrained assessment trials but the total number of options generated did not (cf. Johnson & Raab, 2003 Raab & Johnson, 2007). From a methodological standpoint, tests aimed at predicting skill using option generation data should first categorize options into task-relevant and –irrelevant. This finding also has implications for training and theory. Training for less-skilled players should focus on increased generation of high-quality options—where they exist in the environment (see Ward et al., 2013)—and reducing the generation of low-quality options (as opposed to the generation of all additional options, cf. Johnson & Raab, 2003). Furthermore, this should be trained under increasingly difficult levels of time pressure. From a theoretical standpoint, this finding offers tentative support for LTWM within the soccer scenarios investigated. Future research should investigate the effectiveness of LTWM-based training interventions (see Ward et

al., 2013)—i.e., those that focus upon the generation of more task-relevant information and less task-irrelevant information—in comparison to other training techniques.

Correlations within each of the time constraint conditions were too weak to offer support for either mechanism. For this reason, the hypotheses regarding the time effect upon the relationship between task-irrelevant options and performance did not receive support. Furthermore, the hypotheses regarding the time effect upon the relationship between task-relevant options and performance, particularly within the moderate-skill group, also did not receive support. I suspect that this is because the correlational hypotheses were underpowered. To increase power, I averaged the results across time conditions. When averaged across both time conditions, the correlational data offer support for part of the hypothesis (*H3*). The number of task-relevant options generated was significantly and positively related to anticipation accuracy. Also, the number of task-irrelevant options generated approached significance in the negative direction to anticipation accuracy, as would be expected by proponents of LTWM theory when making decisions in this type of dynamic sporting situation (Ward et al., 2013). The total number of options generated was never negatively related to anticipation accuracy. Although this data appear to provide some very tentative support the use of LTWM-type mechanisms within the assessment-phase of the representative soccer tasks employed during this study—as opposed to the use of simpler heuristics (Ward et al., 2013)—more evidence is needed because of the increased susceptibility to type one error when conducting additional comparisons (i.e., by also averaging data across conditions).

Next, during intervention trials I hypothesized that the moderate-skill group would generate fewer total options, as per TTF, but also explored whether the moderate-skill group would generate more task-relevant and fewer task-irrelevant options than the low-skill group, particularly during time constrained trials (*H4*). The 3-way interaction was not observed, however, the simple interaction effect data were similar to those observed during assessment. The moderate-skill group generated more task-relevant and fewer task-irrelevant options than the low-skill group during time constrained trials only. As for assessment, the data observed during intervention suggest that the moderate-skill group responded to time constraint by reducing the generation task-irrelevant information only, whereas the low-skill group generated less information in general. Again, this is in line with Raab and Johnson (2007)—as one acquires more skill, they are better able to generate a satisficing option early in the generation sequence and they reduce the generation of unnecessary information. However, the interaction effects again reveal tentative support for the use of LTWM-type mechanisms.

As in assessment phase, the presence of a Skill x Information Type interaction effect suggests that numbers of categorized task-relevant and task-irrelevant options were more discriminating of skill group than simply the total number of options generated (see Ward et al., 2013, cf. Johnson & Raab, 2003). This was at least the case during time constrained trials. The utility of categorization of options into task-relevant/task-irrelevant offers tentative support for the use of LTWM-type mechanisms during the intervention phase of decision making, in addition to assessment. Just as in assessment

trials and the work by Ward and colleagues (2013), the generation of task-relevant and reduction in generation of task-irrelevant information differentiated between skill groups more effectively than simply the total number of options (cf. Raab & Johnson, 2007).

Recall that I hypothesized that the total number of options would be negatively related to performance during intervention trials (*H5*). I expected that this would be more prominent during time constrained trials. This was not the case. The total number of options was not related to performance within either time condition. Even when averaged across both time conditions in order to increase power, the TTF-based hypothesis did not receive support. However, I also explored the claims of Ward et al. (2013) within intervention trials. When averaged across time conditions, the numbers of task-relevant and task-irrelevant options were more related to performance (i.e., selection of the best option) than simply the total number of options. This suggests that LTWM-type mechanisms may be employed during the intervention-phase of decision making in addition to the assessment-phase, at least within the soccer scenarios investigated in this research. This finding should be taken very tentatively, given that the relationships were observed across time condition and skill groups which may contribute to the potential for type I error caused by additional analyses. Further evidence in future research is needed to support this claim.

I also speculated that time constraint would affect the negative relationship between task-irrelevant options and performance (*H5*). This relationship was more negative during time constrained trials than during non-time constrained trials, as

hypothesized. Additionally, I expected that the number of task-relevant options and performance would remain unaffected by time constraint. This hypothesis did not receive support because the relationships involving task-relevant options were non-significant. These two findings may contain implications for training. In general during intervention trials, participants who performed well generated fewer task-irrelevant options. Furthermore, high-performing participants tended to adapt to time pressure by generating even fewer task-irrelevant options. This was considerably more profound than generating *more* task-relevant options. By this logic, intervention-phase training in situations of high time pressure (e.g., within the domain of soccer) might benefit from focusing more heavily upon the reduction in generation of task-irrelevant information than it should on an increase in the generation of task-relevant information. In other words, a training intervention in soccer which focuses on ignoring distracting information/options while a player is in possession of the ball may be more effective than focusing attending to all of the high-quality options available. Future research should evaluate the effectiveness of new training techniques (i.e., generating more task-relevant information, generating less task-irrelevant information) in comparison with current training methods used in sport-based decision making.

In the context of the representative soccer task used, Experiment 1 provided some very tentative support for the hypotheses proposed by Ward et al. (2013), which were based on the mechanisms described by LTWM theory (Ericsson & Kintsch, 1995), during both the assessment and intervention phase of decision making. However, there are some

consistencies with TTF (Raab & Johnson, 2007)—the reduction of unnecessary information in response to time constraint was indicative of a level of higher skill.

Experiment 1 also provided initial insight into the effect of time constraint on option generation. While time constraint sometimes affected the option generation *process* of skilled and less-skilled decision makers differently (see Gobet & Simon, 1996a; Calderwood et al., 1988), it did not affect the *performance* of skill groups differently (see Chabris & Hearst, 2003; Holding, 1989; Dunne, 1995). A more conservative time constraint is implemented in Experiment 2 to provide further insight into the effect of time constraint on assessment and intervention phases of decision making.

## **Experiment 2<sup>2</sup>**

Experiment 2 was designed to provide further insight into the relationship between option generation behavior, performance, and the manipulation of time constraint while completing the option generation paradigm in soccer. The results from Experiment 1 provide very tentative evidence in favor of the cognitive mechanisms described by LTWM theory (Ericsson & Kintsch, 1995) may be employed when generating options during both assessment and intervention phases of decision making, specifically in the types of dynamic and uncertain soccer scenarios used in Experiment 1.

---

<sup>2</sup> Experiment 2 was also submitted for publication under Belling, Suss, and Ward (2013; under review B; see reference list for more details). Patrick Belling collected and analyzed data and wrote reports of findings. Dr. Joel Suss assisted with writing. Dr. Paul Ward assisted with analyses and writing.

However, the results did not offer conclusive evidence on the effect of time constraint on decision making during different phases.

To investigate this further, a more conservative time limit was used (i.e., 10s) reducing the time available to respond by 33%. During Experiment 1, participants typically completed the option generation task in around 15s, regardless of the application of time constraint. In general, however, participants took longer than 10s. For this reason, I reduced the time constraint to 10s. My goals were to examine the effects of time constraint on performance and process during assessment and intervention in order to determine if participants used different strategies when under a more severe time constraint. Recall that the results of Experiment 1 did not provide the expected support for TTF when making intervention decisions. Since I used female participants in Experiment 1, I recruited mostly male participants during Experiment 2, following Johnson and Raab (2003). I also focused on a single low-skill group in Experiment 2 and so hypotheses related to skill-based differences were excluded. I examine these questions using a more skilled group in Experiment 3.

### **Hypotheses about Performance during Assessment and Intervention**

*H1.* I expected that performance, as assessed by anticipation accuracy during assessment and selection of the criterion best option during intervention, would be significantly lower during time constrained trials compared to trials that were not time constrained. I also expected lower scores on additional measures of performance during

time constrained, as opposed to non-time constrained trials. Additional performance measures included the generation, accurate rating, and anticipation of the criterion most threatening option during assessment trials and the generation and accurate rating of the criterion best option during intervention trials.

### **Hypotheses about Process during Assessment**

*H2.* In order to explore the effect of time pressure on this task, I observed the number of task-relevant/task-irrelevant/total options generated with and without time constraint present. I speculated that a selective reduction in information may occur. In other words, time pressure would reduce the number of task-irrelevant options, but not the number of task-relevant options. Although the low-skill group in Experiment 1 did not respond to the more lenient (i.e., 15s) time constraint (in Exp. 1) in this way (or at least not as strongly as the moderate-skill group), I speculated that a more conservative time constraint might yield this effect. A selective reduction in information (even within a low-skill group) is in accordance with the expectations of TTF (Johnson & Raab, 2003). Accordingly, I speculated that I would observe a 2-way interaction of information type (within participants: task-relevant/task-irrelevant) by time constraint (within participants: present/absent) during assessment trials.

*H3.* In accordance with Ward et al. (2013), during assessment trials I expected that the number of task-relevant options generated would be positively, and the number of task-irrelevant options will be negatively, related to performance (e.g., anticipation



accuracy of the actual outcome). Just as in Experiment 1, I expected that these relationships would be affected by time constraint. More specifically, I expected that the negative relationship between task-irrelevant options generated and performance would be stronger—in the negative direction—during time constrained trials. This is because I speculate that the generation of irrelevant information would be more detrimental to performance under time pressure than when given unlimited time. Additionally, to explore the claims of TTF, I also observed the relationship between the total number of options and performance within each of the time conditions of assessment trials. If a negative relationship was observed between the total number of options generated and performance, support for TTF would be provided. If task-relevant and task-irrelevant options generated were more related to performance, support of LTWM-type mechanisms would be provided.

### **Hypotheses about Process during Intervention**

*H4.* I recorded the number of task-relevant/task-irrelevant/total options generated during time constrained and non-time constrained trials in order to explore the effects of time constraint. During intervention trials and in agreement with models based on activation by association (see Johnson & Raab, 2003), I speculated that the application of time constraint would reduce the number of task-irrelevant options generated, but not necessarily the number of task-relevant options generated. Accordingly, I speculated that I would observe a 2-way Information Type (within participants: task-relevant/task-

irrelevant) x Time Constraint (within participants: present/absent) interaction effect during intervention trials.

**H5.** In accordance with Johnson and Raab (2003), I expected that the total number of options would be negatively related to performance (i.e., selection of the criterion best option) Similar effects were expected for the additional measures of performance (i.e., generation and accurate rating of criterion best option). I expected the trends associated with TTF (i.e., a negative relationship between the total number of options and performance) would be more prominent during time constrained intervention trials—given the simple heuristic’s emphasis on immediately generating a satisficing option. In other words, I expected that the more conservative time limit during intervention trials would force a shift toward the relationships predicted by TTF—as I hypothesized in Experiment 1. However, in light of these results, and to explore the claims of Ward et al. (2013), I also observed the relationship between task-relevant/task-irrelevant options and performance within each of the time constraint conditions during intervention trials. If the number of task-relevant/task-irrelevant options was positively/negatively related to performance, support for LTWM-type mechanisms would be provided. As in Experiment 1, I expected that generating irrelevant information would degrade performance more during time constrained trials than during non-time constrained trials. Accordingly, I expected that the relationship between number of task-irrelevant options generated and performance would be stronger—in the negative direction—during time time constrained trials than during non-time constrained trials.

## **Methods**

### **Participants**

Twenty-one (17 males) recreational-level players with an average age of 19.8 ( $SD = 1.94$ ) participated in Experiment 2. The participants averaged 10.19 ( $SD = 6.04$ ) years of experience playing under the supervision of a coach. This included any recreational level or pre-collegiate level play. None of the participants in Experiment 2 played soccer at the collegiate varsity level. Participants received course credit for their participation in the research.

### **Materials and Procedure**

Materials included in Experiment 2 were identical to those of Experiment 1 (see above). Participants followed the exact same procedure described above in Experiment 1 with two exceptions. Firstly, in half (i.e., 12) of all assessment and intervention trials (i.e., when time constraint was applied), participants were only allotted 10 seconds, as opposed to 15 seconds, during which they could generate options. Time to rate options was unlimited. In the other half (i.e., 12) of all trials time to generate and rate options was unlimited. Secondly, participants completed an online version of the soccer experience questionnaire in a follow up survey, instead of taking this survey during the option generation paradigm.

## Data Analysis

**Coding and Measures.** The same coding structure and measures were used as in Experiment 1.

**Analysis.** Paired *t*-tests were used to test for effects of time constraint on performance (*H1*).

To test for a 2-way interaction of effect of information type and time constraint on option generation behavior, a factorial ANOVA was used (*H2*, *H4*). Information type (task-relevant/task-irrelevant) and time constraint (present/absent) were the within-participant variables. The number of options generated was the dependent variable. Within-participants effect sizes were calculated to measure the effect of time constraint on option generation. Interaction effects were estimated using partial eta squared.

To test for the relationships between option-generation behavior and measures of performance during assessment (*H3*), I used Pearson's *r*. If the positive/negative relationships between task-relevant/task-irrelevant options, respectively, were stronger (i.e., further from zero) than support for my hypothesis and the LTWM-based claims of Ward et al. (2013) would be provided. Further support for *H3* would be provided if the relationship between task-irrelevant options and performance was stronger during time constrained trials than during non-time constrained trials.

I also used Pearson's *r* to observe the expected TTF claim of a negative correlation between the total number of options and performance during intervention (*H5*). If the total number of options generated was negatively related to performance, and

if this relationship was stronger (i.e., more negative) during time constrained than during non-time constrained intervention trials, support for the TTF-based portion of *H5* would be provided. To explore the possibility that different relationships may be observed for task-relevant and task-irrelevant information during intervention (as in assessment; see Ward et al., 2013), I conducted additional correlations with these variables and performance. If the relationship between task-irrelevant options and performance was more negative during time constrained trials than during non-time constrained trials, support for this portion of *H5* would be provided.

Lastly, a manipulation check was performed using paired *t*-tests. Time constraint was the within-participant factor. The dependent variables were anxiety and effort, as assessed by the MRF-L and RSME, respectively, during both assessment and intervention. Within-participants effect sizes of time constraint were calculated for anxiety and effort.

## **Results**

### **Manipulation Check**

During assessment, participants reported significantly higher levels of anxiety during time constrained trials than during non-time constrained trials,  $t(1, 20) = 2.30, p = 0.03, ES = 0.32$ . However, no differences were observed between time constraint conditions in subjective levels of effort,  $t(1, 20) = 1.38, p = 0.18, ES = 0.14$ . For a

full review of the effect of time constraint on anxiety and mental effort during assessment trials, see Table 6.

**Table 6.** Anxiety and mental effort ratings during time constrained and non-time constrained assessment trials of Experiment 2.

Time condition	Anxiety rating during assessment		Mental effort rating during assessment	
	Time constraint	No time constraint	Time constraint	No time constraint
Mean (SD)	14.29 (6.10)	12.57 (5.41)	63.10 (21.71)	60.00 (22.19)
Effect of time constraint	$t(1, 20) = 2.30, p = 0.03, ES = 0.30$		$t(1, 20) = 3.23, p = 0.08, ES = 0.26$	

During intervention, participants did not report significantly higher levels of anxiety,  $t(1, 20) = 0.85, p = 0.41, ES = 0.11$ , or effort,  $t(1, 20) = -0.18, p = 0.86, ES = -0.02$ , during time constrained compared to non-time constrained trials. For a full review of the effect of time constraint on anxiety and mental effort ratings during intervention trials, see Table 7.

**Table 7.** Anxiety and mental effort ratings during time constrained and non-time constrained intervention trials of Experiment 2.

Time condition	Anxiety rating during intervention		Mental effort rating during intervention	
	Time constraint	No time constraint	Time constraint	No time constraint
Mean (SD)	12.28 (5.25)	11.71 (5.46)	60.48 (21.03)	60.86 (19.78)
Effect of time constraint	$t(1, 20) = 0.85$ $p = 0.41, ES = 0.11$		$t(1, 20) = -0.18$ $p = 0.86, ES = -0.02$	

## Performance during Assessment and Intervention

**H1.** I hypothesized that performance would degrade as a result of time pressure. During assessment and counter to the stated hypothesis, participants did not demonstrate lower anticipation accuracy during time constrained trials than during non-time constrained trials,  $t(1, 20) = -0.22, p = 0.83$  (see Table 8). Similarly, no effect of time constraint on performance was observed during intervention. Participants selected the criterion best option equally often during time constrained and non-time constrained trials,  $t(1, 20) = 1.34, p = 0.20$  (see Table 8). None of the additional measures of performance were significantly affected by time constraint during assessment trials (see Table 15 in appendix) or intervention trials (see Table 16 in appendix).

**Table 8.** Mean (SD) frequency scores for anticipation accuracy during assessment and selection of the criterion best option during intervention and effects of time constraint for Experiment 2.

	Anticipation accuracy of actual outcome	Selection of criterion best option
Time constraint	1.33 (1.28)	1.33 (1.15)
No time constraint	1.43 (1.12)	0.90 (1.14)
Time constraint effect	$t(1, 20) = -0.22$ $p = 0.83, ES = -0.09$	$t(1, 20) = 1.34$ $p = 0.20, ES = 0.38$

## Option Generation during Assessment

**H2.** Counter to the hypothesis, the Information Type x Time Constraint interaction effect was not observed during assessment,  $F(1, 20) = 0.81, p = 0.38, \eta p^2 = 0.04$ . Participants reduced the generation of both task-relevant ( $ES = -0.07$ ) and task-

irrelevant ( $ES = -0.31$ ) information when time constraint was implemented. Accordingly, only a main effect of time constraint on total number of options was observed,  $F(1, 20) = 5.10, p = 0.04, \eta^2 = 0.20$  (for a full review, see Table 9).

**Table 9.** Mean (SD) number and type of options generated, main effect of time, and interaction effect of Time x Information Type during assessment trials of Experiment 2.

	Condition of time constraint	Task-relevant options	Task-irrelevant options
Number of options generated	Time constraint	0.76 (0.33)	0.94 (0.44)
	No time constraint	0.79 (0.44)	1.13 (0.61)
Time constraint effect		$F(1, 20) = 5.10$ $p = 0.04, \eta^2 = 0.20$	
Information Type x Time Constraint interaction effect		$F(1, 20) = 0.81$ $p = 0.38, \eta^2 = 0.04$	

**H3.** During time constrained trials, the relationship between the number of task-relevant options generated and anticipation of the actual outcome approached significance in the hypothesized direction ( $r = 0.32, p = 0.16$ ). However, the number of task-irrelevant options generated was not negatively related to anticipation accuracy ( $r = 0.14, p = 0.54$ ). For a full review of option generation and performance during time constrained assessment trials, see Table 17 in the appendix.

During non-time constrained trials, although the trends were in the expected direction, there was no significant relationship between the number of task-relevant options generated and anticipation accuracy ( $r = 0.28, p = 0.21$ ), or between the number of task-irrelevant options generated and anticipation accuracy ( $r = -0.30, p = 0.19$ ). I



hypothesized that the relationship between task-irrelevant options and performance would be more negative during time constrained trials than during non-time constrained trials. This hypothesis was not supported. In fact, the relationship between task-irrelevant options and performance was more negative during non-time constrained trials ( $r = -0.30$ ,  $p = 0.19$ ) than during time constrained trials ( $r = 0.14$ ,  $p = 0.54$ ). For a full review of relationships during non-time constrained assessment trials, including additional performance metrics, see Table 18 in the appendix.

#### Option Generation during Intervention

**H4.** As hypothesized, an Information Type x Time Constraint interaction was observed,  $F(1, 20) = 5.10$ ,  $p = 0.04$ ,  $\eta p^2 = 0.20$ . During intervention trials, participants reduced the generation of task-irrelevant information ( $ES = -0.70$ ) significantly more than task-relevant information ( $ES = 0.17$ ) when time constraint was implemented. A main effect of time constraint on total number of options generated was also observed,  $F(1, 20) = 5.70$ ,  $p = 0.03$ ,  $\eta p^2 = 0.22$  (for a full review, see Table 10).

**Table 10.** Mean (SD) number and type of options generated, main effect of time, and interaction effect of time by information type during intervention trials of Experiment 2.

	Task-relevant options	Task-irrelevant options
Number of options generated	0.80 (0.41)	0.89 (0.41)
	0.74 (0.35)	1.21 (0.46)
Time constraint effect	$F(1, 20) = 5.70$ $p = 0.03$ , $\eta p^2 = 0.22$	

Information Type x Time Constraint interaction effect	$F(1, 20) = 5.10$ $p = 0.04, \eta p^2 = 0.20$
----------------------------------------------------------	--------------------------------------------------

**H5.** I expected that the total number of options would be negatively related to performance during intervention (see Johnson & Raab, 2003). More specifically, during time constrained trials, I expected a shift toward the use of a stronger TTF-based strategy. Counter to this hypothesis during time constrained trials, the total number of options generated was not negatively related to selecting the criterion best option ( $r = 0.28, p = 0.21$ ).

Recall that my intention was also to explore the predictions of Ward et al. (2013) during time constrained trials. The number of task-relevant options generated was significantly and positively related to selecting the criterion best option when under time constraint ( $r = 0.63, p < 0.01$ ). However, the number of task-irrelevant options generated was not significantly negatively related to the selecting the criterion best option ( $r = -0.27, p = 0.23$ ). For a full review of option generation and performance during time constrained intervention trials, see Table 19 in the appendix.

During non-time constrained trials, the relationship between the total number of options and selecting the criterion best option was not significant ( $r = -0.20, p = 0.38$ ). Recall that I expected a shift toward TTF in the presence of time constraint. Instead, the correlation trended in the direction predicted by TTF (i.e., negative) during non-time constrained trials only. Total options also trended negatively with additional performance

measures. For a full review of option generation and performance during non-time constrained trials, see Table 20 in the appendix.

Also during non-time constrained trials, I explored the LTWM-based claims of Ward et al. (2013). The number of task-relevant options was not related to selecting the criterion best option ( $r < 0.01$ ,  $p = 0.99$ ). However, the relationship between the number of task-irrelevant options and the selection of the criterion best option approached significance in the negative direction ( $r = -0.33$ ,  $p = 0.15$ ). The number of task-irrelevant options generated was also more related to the additional performance measures than task-relevant or total options during non-time constrained trials (see Table 20 in appendix).

## **Discussion**

Counter to the hypothesis (*H1*), the implementation of a more conservative time constraint did not degrade performance during assessment or intervention (cf. Gobet & Simon, 1996a; Calderwood et al., 1988). Just as in Experiment 1, performance was not affected by the implementation of time constraint, but the data suggest that the process/strategy selection was (at least during intervention trials; see below). Again, this suggests that the mechanisms relied upon during this representative task in soccer may differ from more traditional domains (e.g., chess)—i.e., the ability to anticipate and decide upon courses of action during a pattern of play presented in the lab is not affected

the implementation of time constraint. In fact, time constrained trials likely represent the natural ecology more accurately.

Counter to the hypothesis (*H2*), a 2-way interaction of information type and time constraint was not observed during assessment trials. In other words, participants did not demonstrate a selective reduction in information type in response to time constraint. Instead, as in Experiment 1, the low-skilled participants used in this experiment reduced the generation of both task-relevant and task-irrelevant information. The correlational data provide insight in this regard.

During time constrained assessment trials, while the expected relationships between task-relevant/task-irrelevant options and performance were not significant, there were often trends with performance measures in the direction hypothesized (*H3*). When time constraint was removed, many of these trends—using primary and additional measures of performance (see Table 18 of appendix)—reached statistical significance. The generation of more task-relevant information and less task-irrelevant information (as opposed to less total information, see Johnson & Raab, 2003) was associated with higher performance. These results provide very tentative evidence supporting the use of mechanisms of the type described by LTWM as an explanation of skilled anticipation in dynamic and uncertain soccer scenarios of the type used in this study (see Ward et al., 2013). Future research should investigate these relationships further, as the trends presented in this experiment were often non-significant.

Furthermore, the results have implications for future research in training. When given more time to access encoded information after occlusion, participants who were able to generate more task-relevant and less task-irrelevant information performed better. In other words, when permitted time, participants who were able to access more task-relevant information (as opposed to less information in total; see Johnson & Raab, 2003) anticipated the outcome of scenarios more accurately. These trends were less prominent when under time constraint. Future research should evaluate the effectiveness of a training program which focuses upon identifying many of the threatening options available to an opposing player with the ball, without focusing on non-threatening options—even in conditions where time pressure is not applied. This type of training intervention should be evaluated in comparison to current methods used to train decision making in soccer.

During intervention trials, the hypothesized Information Type x Time Constraint interaction was observed (*H4*). Participants engaged in a selective reduction of information such that their response to the implementation of time constraint was to generate less task-irrelevant information, but not less task-relevant information. This might suggest that high-quality options were generated first and subsequent options were lower in quality (see Johnson & Raab, 2003). This data is in agreement with all models of activation by association (e.g., TTF, LTWM). However, the correlational data provide insight into when these models may be employed during this soccer decision making task.

During intervention trials and counter to the hypothesis and the claims of TTF, there was no significant relationship between the total number of options generated and performance, though it trended weakly in the hypothesized direction during non-time constrained trials (*H5*). Even during this condition, a stronger negative relationship was observed between performance and the number of task-irrelevant options. Furthermore, during time constrained trials, the number of task-relevant options generated was positively related to performance. Recall that I hypothesized that time constraint would cause a shift towards the relationships predicted by TTF. Instead, a non-significant shift toward the direction hypothesized by Ward et al. (2013) was observed. Across all time constraint conditions, these data tentatively support the claims of Ward et al. (2013)—based on LTWM theory—more than the TTF heuristic, i.e., a reduction in the generation of task-irrelevant (as opposed to total) options is related to higher performance. Furthermore, the generation of more task-relevant (as opposed to less in total) options was, typically, associated with higher performance.

Again, training implications are present. Counter to the claims of TTF, intervention-based training should be focused on reducing the generation of task-irrelevant information only, not simply reducing the generation of all information to one single intuitive response option. Furthermore, the positive relationship between the number of task-relevant options generated and performance during time constrained trials suggests that training should encourage the rapid generation of *more* task-relevant information. A rapid intuitive response containing multiple (e.g., 2) task-relevant options

was indicative of higher performance, but generating additional options (in total) when permitted more time was not. I speculate that this may be because an intuitive response containing relevant information is representative of a different and more successful strategy than an extended search of the mental representation in order to generate task-relevant (or task-irrelevant) information, though this is outside the scope of this work and further empirical work would be needed to investigate this claim.

These results offer a novel insight into the effect of time constraint on option generation behavior and initial implications for training. Additionally, the results of Experiment 2 reconcile the differences of two lines of research. When participants (in this case a mostly male sample) were asked to generate intervention options during trials with no time constraint present—which is very similar to the methods used by proponents of TTF (Johnson & Raab, 2003; Raab and Johnson, 2007)—the results trended, albeit very weakly, with the predictions of TTF. Perhaps with higher power, a significant TTF-type relationship would have been observed. This offers explanation as to how different findings have been observed across multiple studies (see Ward et al., 2013; Raab & Johnson, 2007). However, in general, the categorization of options into task-relevant/task-irrelevant was useful because the generation of specific information types (i.e., relevant/irrelevant) were more related to performance than simply the total number of options.

I acknowledge limitations in this research. First of all, many of the hypothesized relationships did not reach statistical significance, even if they trended as hypothesized or

in other theoretically interesting directions. I expect that this may be the result of utilizing a low-skill sample. If future research compared data from a recreational-level sample with a higher skill level (e.g., NCAA Division I), the observed trends would provide more detailed insight into the cognitive mechanisms being employed by experts.

### **Experiment 3<sup>3</sup>**

In Experiment 3, my aim was to apply the more stringent time constraint (from Experiment 2) to the option generation task, but also in the context of multiple skill groups (as in Experiment 1). In Experiment 1, there was limited separation between skill groups in terms of performance. Accordingly, the observed differences in process were weaker than expected. In Experiment 3, I tested using the option generation paradigm among NCAA Division 1 soccer players and recreational-level soccer players. My aim was to test the claims of LTWM (i.e., Ward et al., 2013) and TTF (i.e., Johnson & Raab, 2003) within assessment- and intervention-phases of decision making. Because I acknowledge that players of higher skill might adhere to predictions of TTF more closely (see Raab & Johnson, 2007), my hypotheses still reflect the claims of TTF where applicable, despite the findings of Experiments 1 and 2, which tended to provide more support for the use of LTWM-type mechanisms than the TTF heuristic. My specific hypotheses are as follows.

---

<sup>3</sup> Experiment 3 has also been submitted for publication under Belling, Suss, and Ward (under review A). Patrick Belling collected and analyzed data and wrote reports of findings. Dr. Joel Suss assisted with writing. Dr. Paul Ward assisted with analyses and writing.



## **Hypotheses about Performance during Assessment and Intervention**

**H1.** Just as in Experiment 1, I expected to observe a main effect of skill and an interaction effect of Time Constraint x Skill on performance during assessment (i.e., anticipation accuracy) and intervention (i.e., selection of the criterion best option).

## **Hypotheses about Process during Assessment**

**H2.** Just as in Experiment 1, I expected to observe a 3-way interaction effect of Time Constraint x Information Type x Skill. Again, I expected the nature of the Information Type x Skill interaction effect (i.e., that the high-skilled participants would generate more task-relevant and fewer task-irrelevant options than less-skilled participants) to differ at each level of time constraint due to the use of a more adaptive strategy by the high-skill group. To explore the claims possibility that a TTF-type heuristic was used during assessment (see Klein & Peio, 1989), I tested for a main effect of skill on number of options generated.

**H3.** In accordance with the claims of LTWM (see Ward et al., 2013) and Experiments 1 and 2, I expected the numbers of task-relevant and task-irrelevant options to be more predictive of performance (i.e., anticipation accuracy) than simply the total number of options generated. I expected this finding across both time settings. Given that Experiment 1 and 2 supported the LTWM-based relationships, I tested the predictive power of task-relevant and task-irrelevant *together* on performance. More specially, I expected these relationships to be in line with LTWM (i.e., the numbers of task-relevant

and task-irrelevant options to be positively and negatively related to performance, respectively). These relationships could then be compared with the TTF claim (i.e., a negative relationship between the number of options generated and performance) during assessment. To explore the effects of time constraint on these relationships, I speculated that the number of task-irrelevant options generated would be more negatively related to performance during time constrained than non-time constrained trials, since the generation of irrelevant information might hinder performance more when under time pressure.

### **Hypotheses about Process during Intervention**

*H4.* During intervention trials, I expected to observe a 3-way interaction effect of Time Constraint x Information Type x Skill on the number of options generated. Again, I expected the nature of the Information Type x Skill interaction effect to differ at each level of time constraint. If a main effect of skill on the number of options generated existed in the absence of an interaction effect with information type (i.e., if the high-skill group generated fewer options in total, regardless of task-relevance), I would take this as support for TTF.

*H5.* In accordance with Experiments 1 and 2, I expected that the numbers of task-relevant and task-irrelevant options generated would be related to performance (i.e., selection of the criterion best option). I expected this finding across both time settings. As in *H3* during assessment and given the support for LTWM in the previous experiments, I

tested the predictive power of task-relevant and task-irrelevant option generation *together* on performance. I expected that together, this option generation behavior would significantly predict performance. More specifically, I expected that these relationships would be in line with the claims of LTWM (i.e., the numbers of task-relevant and task-irrelevant options to be positively and negatively related to performance, respectively). These relationships could then be compared with the relationship predicted by TTF (i.e., a negative relationship between total options generated and performance). Again, to explore the effect of time constraint on these relationships, I observed the relationship between the total number of options generated and performance within each time setting. Because the implementation of time pressure could cause a shift toward a simpler intuitive heuristic like TTF, I speculated that this relationship might be stronger—in the negative direction—during time constrained trials than non-time constrained trials.

## **Methods**

### **Participants**

The high-skill group was composed of 19 male NCAA Division I soccer players. This skill group averaged 19.78 ( $SD = 1.56$ ) years of age and 11.56 ( $SD = 3.36$ ) years of experience playing soccer. The low-skill group was composed of the 17 male recreational-level soccer players from Experiment 2. The low-skilled group averaged 20.24 ( $SD = 1.86$ ) years of age and 8.06 ( $SD = 6.13$ ) years of experience playing soccer. Only data from the male participants in Experiment 2 were used so that the skill groups

were comparable by gender. The high-skill group was paid USD \$20.00 for completing the experimental tasks; the low-skill group received university course credit.

### **Materials, Procedure, and Coding**

The materials, procedure, and coding used were identical to Experiment 2 (see *Methods* of Experiment 2 above).

### **Analysis**

**H1.** To test for a main effect of skill and a Skill x Time Constraint interaction effect, a factorial ANOVA was employed. Skill (high/low) was a between-subjects variable. Time Constraint (present/absent) was a within-subjects variable. To assess performance, anticipation accuracy was used as the dependent variable during assessment trials and selection of the criterion best option was used during intervention trials (see *Measures* of Experiment 1 above for further detail).

**H2 and H4.** To test for a Skill x Time Constraint x Information Type interaction effect on the number of options generated. Skill (high/low) was a between-subjects variable. Time constraint (present/absent) and information type (task-relevant/task-irrelevant) were within-subject variables. The number of options generated was the dependent variable. To follow up, I used simple interaction effects to examine the effect of information type and skill on the number of options generated within each time constraint setting. Partial eta squared and Cohen's *d* were used to estimate effect sizes.

**H3 and H5.** Given that I expected option generation and performance to be related as in the previous experiments, I now wanted to test the predictive power of task-relevant and task-irrelevant option generation *together* on performance. Accordingly, a multiple regression analysis was conducted so that I could observe the relative contribution of relevant and irrelevant option generation on performance. The numbers of task-relevant and task-irrelevant options generated were used as predictor variables. During assessment, the predicted variable was anticipation accuracy. During intervention, the predicted variable was selection of the criterion best option. To observe the relationships in the same manner as the previous experiments and to display them in the most simple way, follow up correlations were calculated using Pearson's  $r$  (as opposed to Beta weights). To prevent type one error given this type of analysis, these correlations were Bonferroni corrected. To observe the relationship between the total number of options generated and performance during assessment and intervention, I calculated Pearson's  $r$ .

**Manipulation Check.** Paired samples t-tests were used as a manipulation check. Scores on the MRF-L (state anxiety) and RSME (mental effort) during time constrained and non-time constrained trials were used as dependent variables. Within participant effect sizes were calculated to estimate the effect time constraint on anxiety and effort.

## Results

### Manipulation Check

During assessment trials, there was not a significant effect of time constraint on anxiety ratings,  $t(35) = 0.26, p = 0.80, ES = 0.03$ . There was a significant effect of time constraint on mental effort ratings,  $t(35) = 2.01, p = 0.05, ES = 0.16$ . During intervention trials, there was not a significant effect of time constraint on anxiety,  $t(35) = 0.98, p = 0.34, ES = 0.04$ , or mental effort ratings,  $t(35) = 0.41, p = 0.68, ES = 0.09$ .

### **Performance during Assessment and Intervention**

*H1.* During the assessment phase, a main effect of skill on performance was observed  $F(1, 34) = 10.21, p = 0.003, \eta p^2 = 0.23$ . The high-skilled group ( $M = 4.21, SD = 1.62$ ) anticipated the outcome of situation accurately on more trials than the less-skilled group ( $M = 2.65, SD = 1.27$ ). However, a main effect of time constraint was not observed,  $F(1, 34) = 0.08, p = 0.79, \eta p^2 = 0.002$ . Counter to my hypothesis, a Skill x Time Constraint interaction effect was also not observed,  $F(1, 34) = 0.006, p = 0.94, \eta p^2 < 0.001$ . Skilled participants performed better than less-skilled participants across all time conditions.

During the intervention phase, I observed a main effect of skill on performance  $F(1, 34) = 10.53, p = 0.003, \eta p^2 = 0.24$ . The high-skilled group ( $M = 4.16, SD = 1.54$ ) selected the criterion best option more often than the less-skilled group ( $M = 2.36, SD = 1.80$ ). Again, a main effect of time constraint was not observed,  $F(1, 34) = 0.55, p = 0.46, \eta p^2 = 0.02$ . Again, counter to our hypothesis, a Skill x Time Constraint interaction effect was not present,  $F(1, 34) = 0.30, p = 0.59, \eta p^2 = 0.01$ .

## Process during Assessment

**H2.** Recall that I expected a three-way interaction of Skill x Time Constraint x Information Type, such that high-skilled participants would generate more task-relevant and fewer task-irrelevant options than less-skilled participants, and this effect would be stronger during time constrained trials—as a result of the use of a more adaptive option generation strategy. This interaction effect was not observed,  $F(1, 34) = 0.56, p = 0.46, \eta p^2 = 0.02$ . However, an Information Type x Skill interaction effect was observed,  $F(1, 34) = 21.57, p < 0.001, \eta p^2 = 0.39$ . Skilled players generated more task-relevant and fewer task-irrelevant options than less-skilled players, regardless of time constraint (see Table 11). An Information Type x Time Constraint interaction effect was not observed,  $F(1, 34) = 0.49, p = 0.49, \eta p^2 = 0.01$ . In other words, participants reduced the generation of both task-relevant and task-irrelevant information in response to time pressure. Recall we expected the nature of the Information Type x Skill interaction effect to change at each level of time constraint. During time constrained trials, the high-skill participants generated more and fewer task-relevant and task-irrelevant options than the low-skill participants, respectively,  $F(1, 34) = 7.40, p = 0.01$  (see Table 11 for effect sizes within time settings). This effect was also observed during non-time constrained trials,  $F(1, 34) = 13.52, p < 0.001$  (see Table 11 for effect sizes within time settings). These effects were not stronger during time constrained trials, counter to my hypothesis (see Table 11). Lastly to explore the validity of the claims of TTF during assessment, we also tested for a

main effect of skill on number of options generated. This effect was not observed,  $F(1, 34) = 0.43, p = 0.52, \eta p^2 = 0.01$ .

**Table 11.** Mean (SD) number and type of options generated by the high- and low-skill groups during assessment-phase trials of Experiment 3.

	Info. type	Skill group	Time constraint	No time constraint	Both time conditions
Number of options generated	Task- relevant	Low	4.35 (1.97)	4.71 (2.44)	9.06 (3.80)
		High	6.11 (2.23)	6.84 (2.03)	12.95 (3.27)
		Cohen's $d$	$d = 0.84$	$d = 0.95$	$d = 1.10$
	Task- irrelevant	Low	5.59 (2.32)	6.88 (3.95)	12.47 (5.41)
		High	3.47 (2.52)	3.79 (2.78)	7.26 (4.20)
		Cohen's $d$	$d = -0.72$	$d = -0.90$	$d = -1.08$

**H3.** The numbers of task-relevant and task-irrelevant options (irrespective of time constraint condition) were used as predictor variables and anticipation accuracy was the predicted variable in the regression model. These analyses revealed that together, the numbers of task-relevant and task-irrelevant options explained a significant amount of the variance in anticipation accuracy,  $R^2 = 0.41, F(2, 33) = 19.65, p < 0.001$ . To investigate the relationships more specifically, we observed the correlations between the numbers of task-relevant/task-irrelevant options generated and anticipation accuracy, and applied Bonferroni corrections where necessary ( $\alpha = 0.025$  level).<sup>4</sup> The numbers of task-relevant options and task-irrelevant options were significantly positively related and negatively

<sup>4</sup> Pearson correlations were used in place of Beta so that the strength of relationship was most comprehensible, especially given the use of Pearson's  $r$  in previous experiments. Because it is more conventional to report Beta weights with multiple regression analysis, they are provided here. The numbers of task-relevant ( $\beta = 0.57, t = 4.17, p < 0.001$ ) and task-irrelevant ( $\beta = -0.20, t = -1.49, p = 0.18$ ) options generated during assessment trials predicted a significant amount of variance in anticipation accuracy.



trended with performance, respectively ( $r = 0.61, p < 0.001$ ;  $r = -0.32, p = 0.03$ , respectively). Both of these relationships were in the hypothesized direction, in agreement with Ward et al. (2013). These relationships correlated or trended in the directions anticipated by LTWM when we analyzed data from the time constrained ( $r = 0.40, p = 0.02$ ;  $r = -0.19, p = 0.28$ , respectively) and non-time constrained ( $r = 0.39, p = 0.02$ ;  $r = -0.38, p = 0.02$ , respectively) conditions separately.

To explore the claims of TTF during assessment, we also examined the relationship between the total number of options generated and performance. A significant negative relationship was not observed, as would be expected by TTF ( $r = 0.12, p = 0.51$ ). This was still the case when we analyzed data from the time constrained ( $r = 0.14, p = 0.41$ ) and non-time constrained ( $r = -0.03, p = 0.87$ ) conditions separately.

### **Process during Intervention**

**H4.** To test the claims of TTF during intervention, we expected to observe an interaction between skill and time, and main effect of skill on number of options generated. Although the interaction was not observed, the skill main effect was,  $F(1, 34) = 6.17, p = 0.02, \eta p^2 = 0.15$ . The high-skill group generated fewer options in total than the low skill group (see Table 12). However, to detect effects of skill and time constraint on the type of information generated—to explore the LTWM claims during intervention—these analyses were conducted in the context of a 3-way factorial ANOVA (Skill x Time Constraint x Information Type). Although the 3-way interaction was not

significant,  $F(1, 34) = 0.83, p = 0.37, \eta p^2 = 0.02$ , the hypothesized Information Type x Skill effect was observed,  $F(1, 34) = 24.20, p < 0.001, \eta p^2 = 0.42$ . Skilled participants generated more task-relevant and fewer task-irrelevant options than less-skilled participants (see Table 12). Moreover, the type of information generated under time constraint affected all participants,  $F(1, 34) = 5.11, p = 0.03, \eta p^2 = 0.13$ . Participants reduced only task-irrelevant options when time constraint was implemented, resulting in a selective rather than a general reduction in options generated. Recall that we also used simple interaction effects to observe the nature of the Information Type x Skill interaction effect within each time constraint setting. During time constrained trials, the effect was observed,  $F(1, 34) = 11.60, p = 0.002$  (see Table 12 for effect sizes within time settings). During non-time constrained trials, this effect was also observed,  $F(1, 34) = 22.00, p < 0.001$  (see Table 12 for effect sizes within time settings). Counter to my hypothesis, this effect was not stronger during time constrained trials (see Table 12).

**Table 12.** Mean (SD) number and type of options generated by the high- and low-skill groups during intervention-phase trials of Experiment 3.

	Info. type	Skill group	Time constraint	No time constraint	Both time conditions
Number of options generated	Task- relevant	Low	4.88 (2.62)	4.18 (1.98)	9.06 (3.93)
		High	5.79 (1.58)	5.63 (1.92)	11.42 (2.63)
		Cohen's <i>d</i>	<i>d</i> = 0.42	<i>d</i> = 0.74	<i>d</i> = 0.71
	Task- irrelevant	Low	5.53 (2.67)	7.29 (2.93)	12.82 (4.63)
		High	2.68 (1.83)	3.58 (2.52)	6.26 (3.53)
		Cohen's <i>d</i>	<i>d</i> = -1.25	<i>d</i> = -1.36	<i>d</i> = -1.59
	Total	Low	10.41 (3.00)	11.82 (2.88)	22.24 (5.14)
		High	8.47 (2.44)	9.32 (2.96)	17.79 (5.01)
		Cohen's <i>d</i>	<i>d</i> = -0.71	<i>d</i> = -0.86	<i>d</i> = -0.88

**H5.** In line with the claims of Ward et al. (2013), we investigated the relationship between task-relevant and task-irrelevant option generation and performance using a multiple regression analysis. The numbers of task-relevant and task-irrelevant options generated significantly predicted performance,  $R^2 = 0.36$ ,  $F(2, 33) = 10.77$ ,  $p < 0.001$ . As in assessment, the variance was partitioned using correlations with Bonferroni corrections ( $\alpha = 0.025$  level).<sup>5</sup> The numbers of task-relevant ( $r = 0.43$ ,  $p = 0.009$ ) and task-irrelevant ( $r = -0.58$ ,  $p < 0.001$ ) options were significantly related to performance in the directions associated with the LTWM-based claims of Ward et al. (2013).

In order to explore the claims of TTF, we observed the correlation between the total number of options generated and performance (i.e., selection of the criterion best option). Across both time settings, the hypothesized relationship was not significant, but trended in line with TTF and our hypothesis ( $r = -0.24$ ,  $p = 0.16$ ). The relationship did not change when we analyzed data from the time conditions separately. Correlations for the data in time constrained ( $r = -0.21$ ,  $p = 0.21$ ) and non-time constrained ( $r = -0.18$ ,  $p = 0.30$ ) conditions were not significant.

## Discussion

As hypothesized, the high-skill group significantly outperformed the low-skill group during assessment and intervention trials of the option generation paradigm. Skill-

---

<sup>5</sup> Just as in assessment trials, Pearson's  $r$  was used to describe the relationship between task-relevant and -irrelevant options. For convention's sake, Beta weights are presented here. The numbers of task-relevant ( $\beta = 0.30$ ,  $t = 2.16$ ,  $p = 0.04$ ) and task-irrelevant ( $\beta = -0.47$ ,  $t = -3.37$ ,  $p = 0.002$ ) options generated during intervention trials predicted a significant amount of variance in selection of the criterion best option.

based differences were found during anticipation/assessment-phase trials (see Abernethy, 1990; Abernethy & Russell, 1987; Ward et al., 2013; for a review, see Mann, Williams, Ward, & Janelle, 2007) and decision making/intervention-phase trials (see Helsen & Starkes, 1999; Raab & Johnson, 2007; Gorman et al., 2013). Counter to our hypothesis, the skill effect did not interact with time constraint (cf. Gobet & Simon, 1996a). I speculate that time pressure did not affect performance as much as during a representative soccer task as it would during a task associated with a less dynamic domain (e.g., chess).

While time pressure did not affect performance as I expected, it did affect the option generation process. During intervention trials, the hypothesized 3-way interaction was not observed, but a selective reduction (i.e., Time Constraint x Information Type interaction effect) in information was observed. When put under time pressure, participants generated fewer task-irrelevant options, but not fewer task-relevant options. This finding suggests that better options are generated earlier in the decision process and additional option generation only facilitates the generation of low quality options—exactly as would be expected by proponents of TTF (Johnson & Raab, 2003). Likewise, consistent with TTF we observed a negatively trending relationship between the total number of options generated and performance during intervention trials (see Johnson & Raab, 2003) and a skill effect on the total number of options generated (i.e., the high-skill group generated fewer total options than the low skill group) (see Raab & Johnson, 2007).

While these results suggest that TTF provides an accurate description of intervention-phase option generation, the factorial ANOVA and multiple regression analyses provided stronger support for the LTWM-based claims of Ward et al. (2013). The interaction effect of Information Type x Skill was much stronger than the skill effect on the total number of options generated. In other words, the high-skill group generated more task-relevant and fewer task-irrelevant options than the low-skill group. Only because they generated far fewer task-irrelevant options did we observe the skill effect on the total number of options. This finding suggests that the using the numbers of task-relevant and task-irrelevant options was superior for *differentiating between skill groups* than using the total number of options (see Ward et al., 2013; cf. Raab & Johnson, 2007). Likewise, the multiple regression analysis revealed that the numbers of task-relevant and task-irrelevant options generated were more predictive of *performance* than simply the total number of options (cf. Raab & Johnson, 2007). Follow up analyses revealed that these relationships were in accordance with the LTWM-based claims of Ward et al. (2013).

During assessment trials, support for TTF/LTWM was less mixed. A selective reduction in information as a result of time pressure was not observed, which does not offer support for the ‘better options are generated earlier’ notion (cf. Klein & Peio, 1989). Factorial ANOVA revealed that the using the numbers of task-relevant and task-irrelevant options was superior for differentiating between skill groups than using only the total number of options (see Ward et al., 2013; cf. Klein & Peio, 1989). Moreover, the

multiple regression analysis revealed that the numbers of task-relevant and task-irrelevant options generated were more predictive of performance than simply the total number of options. This finding suggests that tools for assessing option generation (presumably to predict skill) should not expect utility from the ‘less-is-more’ approach. During assessment-phase decision making (e.g., defensive tasks in soccer) generating fewer options in total is not indicative of higher skill (cf. Raab & Johnson, 2007) or higher test performance (cf. Johnson & Raab, 2003).

In summary, my findings offer support for the use of LTWM-based mechanisms during the assessment- and intervention-phase decision tasks employed in this research. While the LTWM-based claims of Ward and colleagues (2013) offered the better *explanation* of skilled option generation and decision making during both decision phases, my results still suggested that TTF offers a useful *prescription*. During the intervention phase—where skilled participants generated fewer options than less-skilled participants, the total number of options generated trended negatively with performance, and additional time only permitted the generation of additional task-irrelevant options—soccer players may well benefit from *taking-the-first* option that comes to mind (see Johnson & Raab, 2003).

These results have implication for training in soccer (and perhaps other complex and dynamic domains, though further research is needed in that regard). During assessment or defensive tasks—when the job is to anticipate the actions of an opponent—training should emphasize the generation of as many task-relevant options as are

available in the environment/situation, meanwhile reducing the generation of task-irrelevant options. During intervention or attacking tasks—when the job is make a decision quickly with the ball at one's feet—training should emphasize a quick and intuitive response that focuses on reducing the generation of additional (and often irrelevant) options, in accordance with TTF. However, my results suggest this intuitive response should contain multiple task-relevant options if available in the environment/situation, as opposed to only containing one single option (cf. Johnson & Raab, 2003).

#### **Experiment 4<sup>6</sup>**

Experiments 1, 2, and 3 demonstrated the utility of option generation methods for discriminating between skill groups and investigating the strategies that support superior performance. However, the option generation task is extremely time consuming and effortful to administer and score. It is therefore unlikely to be adopted as a diagnostic tool by professionals (e.g., players, coaches) working in the field. In Experiment 4, my goal was to develop a more practical tool for assessing perceptual-cognitive skill in soccer: the Online Assessment of Skill In Soccer (OASSIS), and examine its validity. OASSIS provides a less time-consuming, easier, and more accessible means of assessing perceptual-cognitive skill, for both the soccer players who complete it and test administrators. Second, OASSIS provides an entirely independent assessment of

---

<sup>6</sup> Experiment 4 has been submitted for publication under Belling and Ward (in press; see reference list). Patrick Belling collected and analyzed data and wrote reports of findings. Dr. Paul Ward assisted with the analyses and writing.

perceptual-cognitive skill in soccer, which can be used to predict the types of strategies used more generally (i.e., option generation behavior recorded on the option generation paradigm). Previously, Ward and colleagues (2013) demonstrated that option generation behavior as recorded on the situational option generation paradigm can be used to predict performance on the same test (albeit during only the assessment-phase trials). In Experiments 1, 2, and 3, I demonstrated this during both assessment- and intervention-phase trials of the updated option generation task. In Experiment 4, an additional goal was to investigate if performance on an *independent* test of anticipation is related to both performance and process (i.e., option generation behavior) during assessment- and intervention-phase trials of the option generation paradigm. This is because I speculate that the same cognitive mechanism (i.e., LTWM) that facilitates skilled option generation also facilitates skilled anticipation and decision making.

OASSIS was created using stimuli that are similar to the option generation paradigm (i.e., video clips of dynamic soccer play that are edited to occlude at critical decision moments), but have been made compatible for use on the computer. While this type of temporal occlusion stimuli has been used in the past and has discriminated between skill groups (e.g., Abernethy, 1990; Abernethy & Russell, 1987; for a review, see Mann et al., 2007), none have been shown to predict performance or option generation behavior *on a separate test* of anticipation or decision making skill. Three main hypotheses for Experiment 4 were as follows:



**H1.** I expected to demonstrate known-groups validity by using performance on OASSIS to demonstrate that high-skilled players perform at a higher level than less-skilled players.

**H2.** Second, I expected to demonstrate predictive validity by observing a positive relationship between anticipation accuracy on OASSIS and anticipation accuracy on the assessment-phase trials of the option generation task. I also speculated that anticipation accuracy on OASSIS would be related to performance during intervention-phase trials of the option generation task (i.e., selection of the criterion best option).

**H3.** Third, I expected to demonstrate convergent validity by testing the predictive power of the numbers of task-relevant/task-irrelevant options on OASSIS performance. I expected that the numbers of task-relevant and task-irrelevant options would predict a significant amount of variance in OASSIS performance. I also expected these relationships to be in the directions expected by LTWM (i.e., OASSIS performance would be positively and negatively related to the number of task-relevant and task-irrelevant options generated). These relationships could then be compared to the relationship expected by TTF (i.e., OASSIS performance would be negatively related to total number of options generated). I expected option generation behavior to predict OASSIS performance in the directions anticipated by LTWM during *both* assessment and intervention trials.

## **Methods**

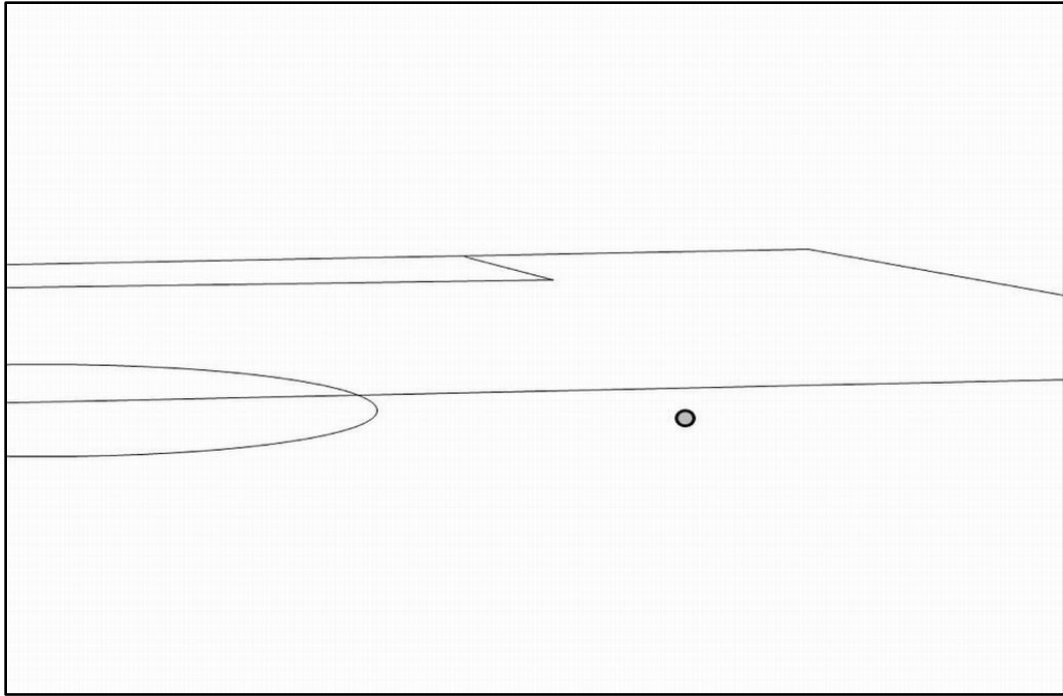
## **Participants**

Fifty male soccer players volunteered to participate and completed OASSIS online. The high-skill group was composed of 13 NCAA Men's Division I players and coaches with an average age of 26.85 ( $SD = 12.84$ ). The low-skill group was composed of 37 recreational-level players with an average age of 20.46 ( $SD = 2.46$ ). Additionally, 19 participants from the low-skill group completed the option generation paradigm. Seventeen of these 19 were the male participants from Experiment 2.

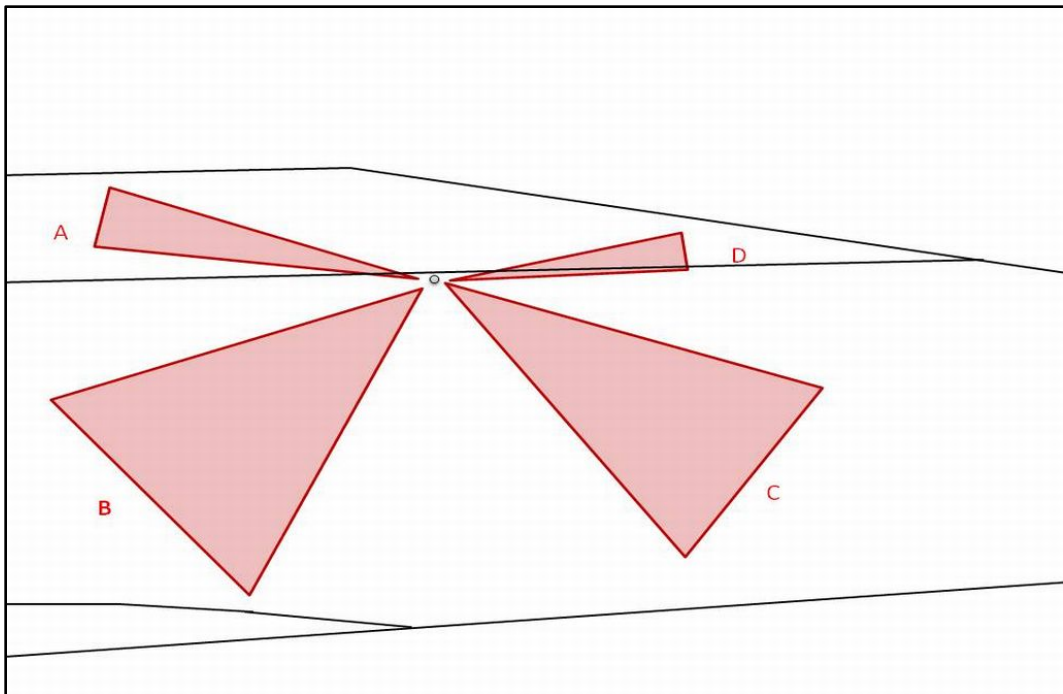
## **Materials**

One hundred and thirty-five video simulations were created using footage from the U.S. Soccer Federation that were similar in length and type to the simulations used in Exp. 1-3. The footage was filmed from a slightly elevated angle above the goal, displaying the length of the pitch and allowing for a pseudo first person perspective (see Mann et al., 2009) and viewing of 11 vs. 11 play (see Ward & Williams, 2003). After 5-10 seconds of dynamic action that conveyed the pattern of buildup play, each clip was edited to end immediately prior to a critical decision moment (i.e., at the moment when the player in possession was about to perform an action with the ball) after which, an occlusion image appeared. This occlusion image matched that of the option generation paradigm, with the exception of the presentation of multiple (e.g., 3-4) options for the participant to choose from, one of which was the actual outcome and, therefore, the correct option to anticipate.

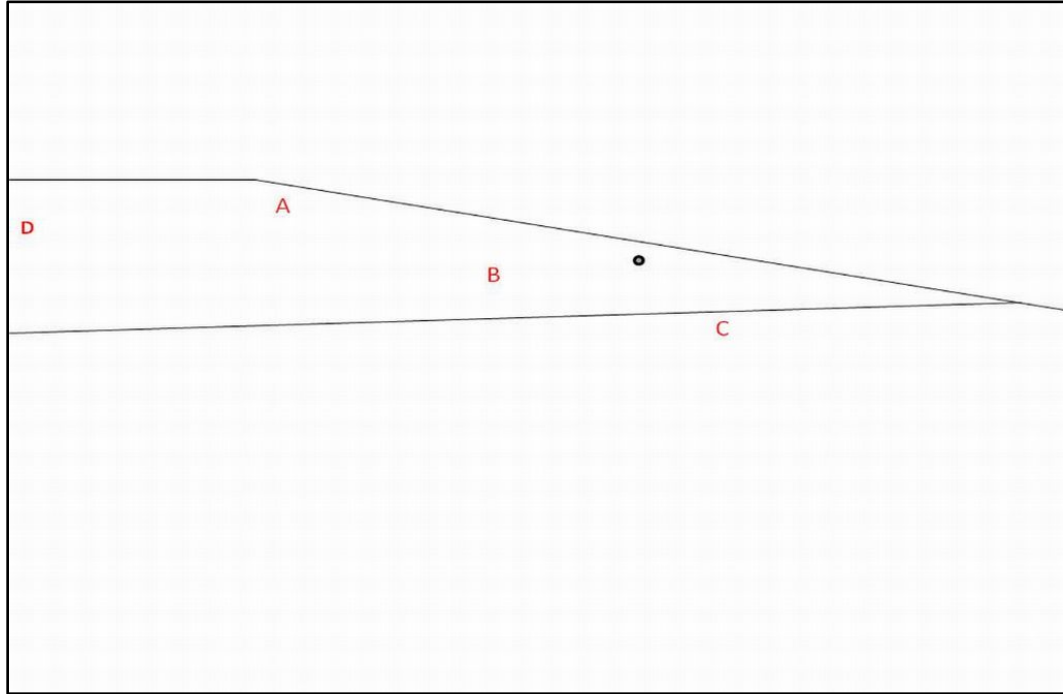
The option generation paradigm requires participants to generate options that include the action(s) of the player with the ball, the direction of play and/or movements of that player, and, in the event of a pass, the location and movement of the recipient and/or point on the pitch where the intended recipient will intercept the pass. In order to best mimic each of these aspects of situational assessment, OASSIS was designed such that each subcomponent could be assessed through multiple-choice testing using three different types of simulations. Action simulations required participants to select the correct action that was about to occur at the time of occlusion from a total three possible actions (e.g., shoot, pass, dribble). Possible responses were presented directly below the video screen (see Figure 4). Direction simulations required participants to select the correct direction in which play was about to continue, regardless of the how (i.e., the action taken) it was going to get there. The correct direction was selected from three or four total directions that were indicated on the video screen. Possible responses were displayed below the video screen and a graphical representation of the possible direction options was added to the line drawing (see Figure 5). Pass-recipient simulations required participants to select the correct recipient of a pass, given the knowledge that the simulation was ending in a pass. The correct recipient was selected from three or four total recipients that were indicated on the video screen and possible responses were displayed below the video screen (see Figure 6). Fifteen simulations for each subcomponent task (e.g., action, direction, and pass-recipient simulations) were created, for a total of 45 unique simulations.



**Figure 3.** The occlusion image for an action simulation of the OASSIS. Response options for pass, shoot, or dribble were displayed below the image.



**Figure 4.** The occlusion image for a direction simulation of the OASSIS. Response options for each direction on-screen were displayed below the image.



**Figure 5.** The occlusion image for a pass-recipient simulation of the OASSIS. Response options for each target player on-screen were displayed below the image.

Following previous research using temporal occlusion methods (Abernethy, 1990; Burroughs, 1984; Abernethy & Russell, 1987; Fadde, 2006; Williams & Ward, 2002), three different versions were created—a pre-contact, at-contact, and post-contact version—for each of the 15 simulations, creating 45 simulations for each simulation type (e.g., 15 x pre-contact action, 15 x at-contact action, 15 x post-contact action). Pre-contact simulations were occluded to end unexpectedly at exactly two frames prior to the

frame in which the on-screen player made contact (e.g., foot-, head-, chest-) with the ball. At-contact simulations were occluded at exactly the frame in which the on-screen player made contact with the ball. Post-contact simulations were occluded at exactly two frames after the frame in which the on-screen player made contact. The response screen—shown after occlusion—was depicted by a line drawing of the same video frame as illustrated in Figures 3-5.

## **Procedure**

Participants were recruited via email to participate in an online test. Participants were directed from the email to the stimuli of the OASSIS using an electronic link. The test was created using Qualtrics.com, a platform for embedding many types of files, including video, into an online survey. A link in the email instantly directed participants to the online consent form. Upon consent, and confirmation that the participant's computer was able to run the online test, participants were trained using one clip of each type (e.g., Action, Direction, and Pass Recipient). After training, the participants were directed to one of three video simulation playlists. Playlists ensured that each participant would respond to only one of the three versions (e.g., pre-contact only) of each of the 15 video simulations in each of the 3 types of simulation (e.g., action, direction, pass recipient). Each of the playlists also randomized the order in which the video simulations were shown. Once the participant completed 45 simulations, they were immediately

directed to an online version of the soccer experience questionnaire used during the option generation paradigm.

## **Data Analysis**

**Coding and Measures.** Data from the OASSIS was coded automatically as the frequency of trials that were answered correctly. If a participant answered all of the trials correctly, they would receive a score of 45. If a participant answered none of the trials correctly, they would receive a score of 0. Data from the option generation paradigm was coded exactly as in Experiments 1 and 2 (see above).

**Analysis.** A one-way ANOVA was used to test for an effect of skill (between participants: high/low) on performance on OASSIS. Cohen's  $d$  was calculated to measure the size of effect of skill on anticipation accuracy during OASSIS trials ( $H1$ ). To test the relationship between performance on the OASSIS and the option generation task, correlational analysis was performed between anticipation accuracy on OASSIS and anticipation accuracy on the option generation task (i.e.,  $H2$ ). To observe the relationships between option generation behavior and performance on the OASSIS, multiple regression analyses were conducted. The numbers of task-relevant and task-irrelevant options generated were used as predictor variables. Anticipation accuracy on OASSIS was used as the predicted variable. Follow up correlations were conducted using Pearson's  $r$  (as opposed to Beta weights) to present the correlations in the simply and in the same manner as previous experiments. To avoid type one error, these correlations

were Bonferroni corrected ( $\alpha = 0.025$ ). To observe the relationship between the total number of options generated and performance, Pearson's  $r$  was used.

## Results

**H1.** A significant skill effect was observed on performance on the OASSIS. The high-skill group ( $M = 32.08$ ,  $SD = 2.78$ ) anticipated significantly more trials (out of 45) correctly than the low-skill group ( $M = 26.08$ ,  $SD = 5.53$ ) on the OASSIS,  $F(1, 48) = 13.88$ ,  $p = .001$ ,  $d = 1.37$ .

**H2.** As hypothesized, anticipation accuracy on the OASSIS was significantly correlated with anticipation accuracy of the actual outcome on the option generation paradigm during assessment trials ( $r = 0.59$ ,  $p = 0.008$ ). However, anticipation accuracy on the OASSIS was not significantly correlated to selection of the criterion best option during intervention ( $r = 0.13$ ,  $p = 0.60$ )

**H3.** When comparing OASSIS performance with option generation during assessment trials, the hypothesized regression equation approached significance,  $R^2 = 0.26$ ,  $F(2, 16) = 2.78$ ,  $p = 0.09$ . To follow up, Pearson's correlations were conducted and Bonferroni corrected.<sup>7</sup> The number of task-relevant options generated approached a significant relationship with OASSIS performance in the hypothesized direction ( $r = 0.48$ ,  $p = 0.04$ ). The number of task-irrelevant options generated was not negatively

---

<sup>7</sup> Beta weights associated with the predictor variables of the regression model are reported here. The number of task-relevant ( $\beta = 0.50$ ,  $t = 2.30$ ,  $p = 0.04$ ) and task-irrelevant ( $\beta = 0.15$ ,  $t = 0.71$ ,  $p = 0.49$ ) options generated during assessment trials did not explain a significant amount of variance in OASSIS performance, though it approached significance.



related to OASSIS performance ( $r = 0.11, p = 0.65$ ), as hypothesized. To investigate the claims of TTF, the relationship between total options generated and anticipation accuracy as observed. This was not negative as would be expected if TTF were being employed during assessment ( $r = 0.39, p = 0.10$ ).

When comparing OASSIS performance with option generation during intervention trials, the hypothesized regression equation approached significance,  $R^2 = 0.30, F(2, 16) = 3.40, p = 0.06$ . Just as in assessment trials to follow up, Pearson's correlations were conducted and Bonerroni corrected.<sup>8</sup> The number of task-relevant options generated was significantly and positively related to OASSIS performance ( $r = 0.54, p = 0.02$ ). The number of task-irrelevant options generated was not negatively related to OASSIS performance ( $r = 0.01, p = 0.96$ ). The number of total options generated not negatively related to performance, as would be expected if TTF was being employed during intervention ( $r = 0.43, p = 0.07$ ).

## Discussion

Full support for *H1* was observed. In fact, the effect of skill was very large. Anticipation has been demonstrated to be highly predictive of skill-based differences in many sport domains (e.g., Abernethy, 1990; Abernethy & Russell, 1987; Williams & Davids, 1995; Helsen & Starkes, 1999; Savelsbergh et al., 2002; Ward & Williams, 2003;

---

<sup>8</sup> The number of task-relevant ( $\beta = 0.55, t = 2.61, p = 0.02$ ) and task-irrelevant ( $\beta = 0.10, t = 0.45, p = 0.66$ ) options generated during intervention trials did not explain a significant amount of variance in OASSIS performance, though it approached significance.

Ward et al., 2013; Rosalie & Müller, 2013). This effect demonstrates the known-groups validity of the OASSIS.

As hypothesized, the predictive validity of the OASSIS was demonstrated by the positive relationship between performance on the OASSIS and performance during assessment trials on the option generation paradigm (*H2*). Surprisingly, performance on OASSIS was not significantly related to performance during intervention-phase trials of the option generation paradigm. Presumably, anticipation accuracy (i.e., assessment-phase performance) and selection of the criterion best option (i.e., intervention-phase performance) are separate skills, and therefore may each differ within and between individuals. By this logic, it would be sensible that anticipation accuracy on the OASSIS is only predictive of anticipation accuracy on the option generation paradigm, though more work is needed to substantiate this claim. I speculate that an intervention-phase version of OASSIS would be more likely to be related to intervention-phase performance on the option generation task, though future research is needed in this regard.

Partial support was observed for *H3*. The regression equations only approached significance. Upon closer look performance on the OASSIS was positively related to the generation of task-relevant options during assessment and intervention (albeit just outside of the  $\alpha = 0.025$  level during assessment), but not negatively related to the generation of task-irrelevant options during assessment or intervention. This finding demonstrates partial support for the convergent validity of anticipation (during the OASSIS) and option generation behavior (during the option generation paradigm). Furthermore, these

relationships provide more support for the claims of Ward et al. (2013) and for the cognitive mechanisms described by LTWM (Ericsson & Kintsch, 1995) than they do for TTF (Johnson & Raab, 2003). If data were collected with a high-skill group using both the OASSIS and the option generation paradigm, I would expect to observe additional support for the claims by Ward et al. (2013) (i.e., a negative relationship between task-irrelevant options generated and performance on the OASSIS).

While the results of Experiment 4 are promising, further work is needed. The correlations between the option generation paradigm and OASSIS were only demonstrated among the 19 players who participated in both tests. These 19 players were of recreational level in skill. Future research requires data collected from both the option generation paradigm and OASSIS with a high-skill group so that the hypothesized relationships can be observed (or not) among experts. This is particularly important considering that the mechanisms investigated in this line of research (e.g., LTWM, TTF) have been proposed to explain expert performance.

### **Experiment 5<sup>9</sup>**

In Experiment 5, I aimed to further demonstrate the validity of OASSIS. First of all, I aimed to replicate the findings of Experiment 4 by demonstrating a skill effect on performance. Secondly, I observed the relationship between OASSIS performance and assessment- and intervention-task performance on the option generation task within a

---

<sup>9</sup> Experiment 5 has also been submitted for publication under Belling, Suss, and Ward (under review A). Patrick Belling collected and analyzed data and wrote reports of findings. Dr. Joel Suss assisted with writing. Dr. Paul Ward assisted with analyses and writing.

group that contained players of high-skill. I also sought to observe the relationship, within this skilled sample, between OASSIS performance and option generation behavior (i.e., numbers of task-relevant and task-irrelevant options generated), as recorded on the option generation task. Specific hypotheses are as follows:

***H1.*** I expected that the high-skill group would outperform the low-skill group in regard to anticipation accuracy on the OASSIS. If this effect was observed, I would take it as further support for the known-groups validity of the OASSIS.

***H2.*** Among all participants (high- and low-skill), I expected that anticipation accuracy on the OASSIS would be positively related to anticipation accuracy on the assessment-phase task of the option generation paradigm. I also speculated that OASSIS performance would be positively related to selection of the criterion best option on the intervention-phase task of the option generation paradigm. If this relationship was observed, I would take it as further support for the predictive validity of the OASSIS.

***H3.*** Among all participants, I expected option generation behavior during assessment trials of the option generation task would be significantly predictive of anticipation accuracy on the OASSIS. Likewise, I expected that option generation behavior during intervention trials of the option generation task would be significantly predictive of anticipation accuracy on the OASSIS. More specifically, I expected the numbers of task-relevant and task-irrelevant options generated (during both assessment and intervention) to be positively and negatively related to OASSIS performance, in the direction expected by LTWM (see Ward et al., 2013). Given that no support was found

for TTF during Experiment 4, I did not compare the LTWM-based relationships with the relationship between total options generated and OASSIS performance.

**H4.** Because anticipation accuracy and option generation behavior during the assessment trials of the option generation task were related to OASSIS performance in Experiment 4, I sought to determine their *relative* contribution to predicting OASSIS performance in Experiment 5. Therefore, a partial mediation table will be constructed. I expect option generation behavior to be significantly predictive of performance on OASSIS (see *H3* above). In turn, I expect that when anticipation accuracy on the option generation paradigm is added to the regression model, the model will explain significantly more variance in OASSIS performance.

## **Methods**

### **Participants**

***Option generation task.*** The data from Experiment 3 was used.

***The OASSIS.*** The participants from Experiment 3 were recruited to complete the OASSIS. High- and low-skill groups were formed based on the level of competition players had reached (see *Participants* of Experiment 3 above).

### **Materials, Procedure, and Scoring**

***Option generation task.*** The results from Experiment 3 were used.

***The OASSIS.*** The same materials, procedure, and scoring from Experiment 4 were used.

### **Analysis and Results**

Before conducting the analyses, we examined the time taken by each participant to complete each test trial to determine whether players had completed the task in accordance with the instructions. Three members of the skilled group were excluded from the data because their average time fell more than 2 standard deviations outside of the mean. In two cases, we speculated that the individuals may have found a way to watch video clips more than once since the time taken was much longer than average. In one case, the time taken was much shorter, and shorter than required to complete the task by watching the entire length of each video clip. The times for the remaining participants were consistent with having watched each video once and responded to each question within a few seconds. Therefore, 16 of the players composing the high-skill group and 17 of the players composing the low-skill group in Experiment 3 were used in the subsequent analyses.

***H1.*** To test for an effect of skill on OASSIS performance, a one way ANOVA was used. As hypothesized, the high-skill group ( $M = 28.06$ ,  $SD = 4.21$ ) anticipated the outcome of the OASSIS trials significantly more often than the low-skill group ( $M = 24.12$ ,  $SD = 4.94$ ),  $F(1, 31) = 6.06$ ,  $p = 0.02$ ,  $d = 0.86$ .

**H2.** To observe the relationship between OASSIS performance and anticipation accuracy on the option generation task, Pearson's correlations were calculated. As hypothesized, performance on the OASSIS was significantly related to anticipation accuracy during assessment-phase trials of the option generation task ( $r = 0.35, p = 0.04$ ). However, performance was not significantly related to selection of the criterion best option ( $r = 0.23, p = 0.20$ ).

**H3.** To compare option generation behavior with OASSIS performance, multiple regression analysis was used. The numbers of task-relevant and task-irrelevant options were entered into the model as predictor variables. Anticipation accuracy on OASSIS was entered as the predicted variable. Follow correlations were calculated using Pearson's  $r$  and were Bonferroni corrected to avoid type one error.

When comparing option generation behavior during assessment with OASSIS performance, the hypothesized regression model was significant,  $R^2 = 0.25, F(2, 30) = 5.03, p = 0.01$ . To follow up, Pearson's correlations were calculated and Bonferroni corrected ( $\alpha = 0.025$ ).<sup>10</sup> The number of task-relevant options was significantly and positively related to OASSIS performance ( $r = 0.49, p = 0.004$ ). The number of task-irrelevant options generated trended in the hypothesized direction with OASSIS performance ( $r = -0.22, p = 0.22$ ).

---

<sup>10</sup> Beta weights for the predictor variables are provided here because it is conventional to report with multiple regression analysis, though Pearson's  $r$  was reported in the results. The number of task-relevant ( $\beta = 0.46, t = 2.85, p = 0.008$ ) and task-irrelevant ( $\beta = -0.11, t = -0.65, p = 0.52$ ) options generated during *assessment* trials of the option generation task explained a significant amount of variance in OASSIS performance.

When comparing option generation behavior with OASSIS performance during intervention, the hypothesized regression model was significant,  $R^2 = 0.29$ ,  $F(2, 30) = 6.02$ ,  $p = 0.006$ . Again, to follow up, Pearson's correlations were calculated and Bonferroni corrected ( $\alpha = 0.025$ ).<sup>11</sup> The number of task-relevant options generated was significantly and positively related to OASSIS performance ( $r = 0.53$ ,  $p = 0.001$ ), as hypothesized. The number of task-irrelevant options generated trended in the hypothesized direction ( $r = -0.22$ ,  $p = 0.21$ ).

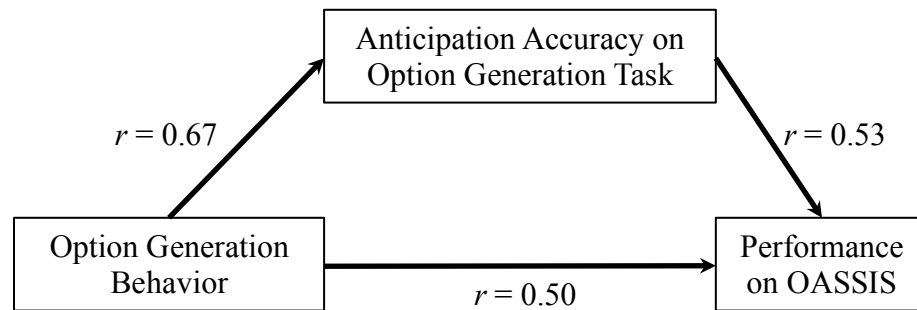
**H4.** Given that anticipation accuracy on the option generation paradigm was related to anticipation accuracy on the OASSIS during Experiment 4, another goal was observe the predictive power of option generation behavior *and* anticipation accuracy on the option generation paradigm, together, for OASSIS performance, hierarchical multiple regression analyses were used. Variance is partitioned here using Beta, so the change in fit (when anticipation accuracy on the option generation task is added to the model) can be most easily interpreted. During assessment, the numbers of task-relevant ( $\beta = 0.46$ ,  $t = 2.85$ ,  $p = 0.008$ ) and task-irrelevant ( $\beta = -0.11$ ,  $t = -0.65$ ,  $p = 0.52$ ) options together predicted anticipation accuracy on the OASSIS,  $R^2 = 0.25$ ,  $F(2, 30) = 5.03$ ,  $p = 0.01$  (as in H3 above). When anticipation accuracy ( $\beta = -0.22$ ,  $t = 1.04$ ,  $p = 0.31$ ) was included as a mediator (i.e., entered in the second step of the regression equation) with the numbers of task-relevant ( $\beta = 0.60$ ,  $t = 2.88$ ,  $p = 0.007$ ) and task-irrelevant ( $\beta = -0.14$ ,  $t = -0.85$ ,  $p = 0.40$ ) options, the regression model did not explain significantly more variance,  $R^2 = 0.28$ ,

---

<sup>11</sup> The number of task-relevant ( $\beta = 0.51$ ,  $t = 3.15$ ,  $p = 0.004$ ) and task-irrelevant ( $\beta = -0.06$ ,  $t = -0.39$ ,  $p = 0.70$ ) options generated during *intervention* trials of the option generation task explained a significant amount of variance in OASSIS performance.



change in  $F(2, 30) = 1.08, p = 0.31$ , though the relationships moved in the hypothesized directions (for a summary, see Figure 6).



**Figure 6.** Mediation table of option generation behavior, anticipation accuracy during assessment-phase trials of the option generation paradigm, and OASSIS performance.

## Discussion

The findings of Experiment 5 provide further support for the OASSIS as a tool for assessing perceptual-cognitive skill. A skill effect on performance was demonstrated, replicating the finding of Experiment 4. Additionally, the hypothesized relationship between performance on the OASSIS and option generation task (at least during assessment-phase tasks) was established within a population that included highly skilled players. As in Experiment 4, I speculate that an intervention-phase version of OASSIS might be more related to performance during the intervention-phase of the option generation task. In this experiment, I have established that performance on the assessment version of OASSIS is at least related to performance on the assessment version of the option generation task. Moreover, the hypothesized relationships between OASSIS

performance and option generation behavior were demonstrated. Interestingly, the number of task-relevant options generated was typically associated with stronger (i.e., further from zero)  $r$  values than the number of task-irrelevant options when being correlated with OASSIS performance, just as in Experiment 4. While these relationships were both in the direction anticipated by Ward et al. (2013), my results suggest that the generation of task-relevant information is more related to (and perhaps more critical to) high performance on this *separate* test of perceptual-cognitive skill (i.e., the OASSIS). Further work is needed in this regard, as this was not the main focus of this experiment.

### Experiment 6<sup>12</sup>

Experiments 4 and 5 demonstrate the utility of the OASSIS for differentiating between skill groups and predicting performance and option generation behavior on a more in-depth test of perceptual-cognitive skill in soccer. In Experiment 6, I aimed to investigate the argument put forth by Voss et al. (2010)—that domain-general cognitive abilities are also predictive of skill. More plainly, while the effects are expected to be small, skilled athletes have tended to have higher general cognitive ability. We examined whether performance on OASSIS could predict skill group membership beyond measures of domain-general cognition, specifically those that tap into the constructs identified as predictive of skill by Voss and colleagues. To assess general cognitive ability, we administered the Berlin Numeracy Test (BNT). The BNT was originally designed as a

---

<sup>12</sup> Experiment 6 has also been submitted for publication under Belling, Suss, and Ward (under review A). Patrick Belling collected and analyzed data and wrote reports of findings. Dr. Joel Suss assisted with writing. Dr. Paul Ward assisted with analyses and writing.

test of statistical literacy and numeracy, but Cokely et al. (2012) demonstrated that it is a robust psychometric measure of domain-general cognition. It is also significantly predictive of more in-depth measures of attention (e.g., Working Memory Span; see Unsworth & Spiller, 2010) and cognitive ability (e.g., Raven's Advanced Matrices Test; see Raven, 2000).

Anticipation and decision making on-the-field in soccer, and also performance on the OASSIS, forces participants to map out players and their actions over space and time—a somewhat effortful mental task. For this reason, it seems plausible that a more general measure of spatial intelligence might also predict soccer skill group membership. Accordingly, we included the Mental Rotations Test (MRT; Vandenburg & Kuse, 1978) as a domain-general cognitive predictor in our analyses and used the updated MRT-A using Qualtrics.com (for specific items, see Peters et al., 1995). This test is a valid measure of general spatial skill (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Consistent with Voss and colleagues, we expected that the BNT and MRT would explain some of the variance in skill level. However, we also predicted that performance on the OASSIS (anticipation accuracy) would predict skill group membership and explain more variance between skill groups than the domain-general measures.

Domain-specific knowledge has also been shown to be predictive of skill and positively related to domain-specific skills (e.g., French & Thomas, 1987; McPherson & French, 1991). In Experiment 6, I aimed to investigate the relative contribution of performance on OASSIS and soccer-specific knowledge. I also aimed to demonstrate

convergent validity for the OASSIS by observing a positive relationship between soccer knowledge and OASSIS performance. To assess soccer knowledge, I created a brief questionnaire in line with previous research on sport-specific knowledge (see Figure 9 and Figure 10 in appendix) (e.g., Hambrick & Engle, 2002).

### **Hypotheses about Performance on the OASSIS and Domain-General Cognition**

*H1.* I expected performance on the OASSIS to be more predictive of skill group membership than the BNT or MRT.

### **Hypotheses about Performance on the OASSIS and Domain-Specific Knowledge**

*H2.* I speculated that performance on the soccer knowledge questionnaire would also predict skill group membership, though not to the extent of performance on the OASSIS.

*H3.* I expected to demonstrate convergent validity for the OASSIS observing a significant positive relationship with the soccer knowledge questionnaire.

## **Methods**

### **Participants**

The low-skilled group was composed of 35 male recreational-level soccer players, 13 of whom also participated in Experiment 2. The high-skill group was composed of the

16 NCAA Division I soccer players from Experiment 2. Participants were recruited via email and on a voluntary basis.

### **Materials and Scoring**

We created online versions of the BNT and MRT. These tests were administered using Qualtrics.com survey builder and then embedded together into one survey.

Participants who had not participated in Experiment 4 or 5 were also directed to the Qualtrics.com website to complete the OASSIS (see Experiment 2). Each of the cognitive assessments and a calculated final score are described below. The OASSIS was scored as in Experiment 4 and 5.

The soccer knowledge questionnaire contained eight questions regarding the rules and typical game-play in soccer (see Figure 9 and Figure 10 in appendix). It was made brief to prevent drop out by participants, given the large battery of cognitive tests being administered in this research. The questions were formatted for multiple choices, though a correct answer sometimes included more than one response option. A final score was calculated out of eight total questions.

The BNT (Cokely et al., 2012) contains four psychometrically validated questions to assess statistical numeracy and risk literacy, which were formatted in Qualtrics.com survey builder (for full review, see Cokely et al., 2012). Participants answered questions by typing a proportion or percentage into a response box. Each correct answer was awarded one point. A final score was calculated out of four points.

An AutoCAD-redrawn version of the original MRT (see Vandenberg & Kuse, 1978) was used in this research. This updated version is referred to as the MRT-A (see Peters et al., 1995). This version contains electronic images that were clearly defined and easily compatible with the Qualtrics.com survey builder. The MRT-A contained 24 items (as in Peters et al., 1995). For each item, a target shape and four response-option shapes were shown to the participant. Two of the response-option shapes were identical, but rotated versions of the target shape. The remaining two response-option shapes were non-identical shapes to the target shape. Participants were required to select the matching (i.e., rotated, but identical) shapes from among the response options. Following Peters et al. (1995), participants completed three training items with feedback before completing the scored items. The 24-item test was divided into two 12-item sets and participants were given three minutes, which was displayed on a timer, to complete each set of 12 items. Participants were awarded one point for each identical shape identified. Given that there were two possible correct answers for each of twenty-four items, a final score out of forty-eight was calculated.

## **Procedure**

Participants received an email containing a link that directed them to the Qualtrics.com survey containing each of the tests. Instructions for each test were provided immediately prior to that test. Participants completed all tests by clicking the mouse to select options and/or typing in responses. During the survey, the “back button”

was disabled, preventing participants from changing responses. When participants finished all the tests, they were thanked for their participation and the survey automatically ended. Participants were not able to ascertain feedback or re-do any test questions, though participants were instructed that they could contact an experimenter for feedback on their scores if they so desired, once their participation was completed.

## **Analysis and Results**

### **Performance on the OASSIS and Domain-General Cognition**

*H1.* A multiple logistic regression analysis was performed using scores from the OASSIS, BNT, and MRT-A to predict skill group membership, which was dichotomous. To compare the relative contribution of the domain-general cognitive tests and OASSIS performance, the Beta weights ( $\beta$ ) were associated with each variable were compared. Performance on OASSIS was a significant predictor of performance ( $\beta = 10.81$ ,  $S.E. = 4.51$ ,  $p = 0.02$ ). The BNT ( $\beta = -0.04$ ,  $S.E. = 0.04$ ,  $p = 0.34$ ) and MRT-A ( $\beta = -0.57$ ,  $S.E. = 0.40$ ,  $p = 0.16$ ) were not significant predictors of skill group membership.

### **Performance on the OASSIS and Domain-Specific Knowledge**

*H2.* When used together to predict skill group membership in a multiple logistic regression analysis, the soccer knowledge questionnaire ( $\beta = 1.12$ ,  $S.E. = 0.39$ ,  $p = 0.01$ ) actually accounted for more variance than performance on OASSIS ( $\beta = -3.04$ ,  $S.E. = 4.75$ ,  $p = 0.52$ ).

**H3.** Performance on OASSIS was significantly and positively related to performance on the soccer knowledge questionnaire ( $r = 0.67, p < 0.001$ ).

## **Discussion**

When compared to domain-general cognitive abilities, performance on the OASSIS was significantly more predictive of skill. In combination with Experiments 4 and 5, this suggests that the OASSIS may be a useful tool for domain professionals (e.g., players and coaches). Furthermore, it strengthens the argument put forth by previous research that domain-specific skills are generally more predictive of skill than domain-general abilities (Helsen & Starkes, 1999; Ward & Williams, 2003; cf. Voss et al., 2010).

However, the utility of OASSIS as an applied tool was put into question by our findings involving domain-specific knowledge. Although domain-specific knowledge and OASSIS performance were strongly related—thereby demonstrating convergent validity for anticipation and sport-specific knowledge—the soccer knowledge questionnaire captured far more between group variance than performance on the OASSIS did. Given this finding, it might seem logical to forego the use of OASSIS or other tools for assessing perceptual-cognitive skill in favor of a sport-specific knowledge assessment. However, I speculate that this may be due to the skill groups used in this research. I recommend that future research employ more than two skill groups. It seems logical that sport-specific knowledge would be extremely useful for differentiating between players who play rarely and players who play very often, as has been demonstrated in this



experiment. However, when attempting to differentiate between moderate- and high-skill groups, where presumably everyone possesses ample knowledge of the sport, I expect that anticipation accuracy (as assessed by the OASSIS) would be more discriminative.

### **General Discussion**

Across all six experiments, this body of work offers contribution to theory and applied science. In Experiments 1, 2, and 3, I tested the claims of two prominent theories, as they pertain to option generation behavior. I tested these claims using an option generation paradigm, which recorded option generation behavior and performance on a representative task. Recall that the claims of LTWM theory (see Ward et al., 2013) are that the numbers of task-relevant and task-irrelevant options generated during such a task are positively and negatively related to performance, respectively. Also, recall that these claims had previously only been applied to the assessment-phase of decision making. In contrast, the claim of TTF (see Johnson & Raab, 2003) is that the total number of options is negatively related to performance. This claim has typically been applied only to the intervention-phase of decision making. For this reason, I tested the claims of these theories using assessment- and intervention-phase tasks on the updated generation paradigm. In general, I expected that the claims of LTWM theory would hold during assessment-phase and that the claims of TTF would hold during intervention-phase. I also anticipated that time pressure may cause a shift toward the relationships predicted by

TTF—given the heuristic’s emphasis on producing a workable solution quickly. In general, my findings were a bit more complex than this.

Experiment 1 provided more support—albeit tentatively—for the claims of LTWM than TTF—across assessment and intervention trials. However, I speculated that two methodological changes might improve the ability of this line of research to test these theoretical claims. First, I implemented a more stringent time limit for the time constrained trials during Experiment 2, since participants often finished comfortably within the time limit during Experiment 1. Secondly, I speculated that the skill groups used in subsequent research should be more defined. This recommendation was applied in Experiment 3.

During Experiment 2, the claims of LTWM held, in general, across assessment-phase trials, regardless of time pressure. During the intervention-phase trials of Experiment 2, the implementation of time constraint appeared to cause a shift *away* from the relationships anticipated by TTF. My results actually trended with TTF only during non-time constrained trials, suggesting that the ability to generate *multiple* task-relevant options was indicative of higher performance. However, this provided an explanation for the seemingly controversial findings across previous research.

Experiment 3 provided the opportunity to test the claims of LTWM and TTF in the context of more stringent time pressure and better defined skill groups. By comparing NCAA Division I and recreational-level players, I was able to determine that the numbers of task-relevant and task-irrelevant options generated was more predictive of skill (i.e.,

the results of the factorial ANOVA) and more predictive of performance (i.e., the results from the multiple regression analysis). This finding held across both assessment- and intervention-phase tasks. Skilled participants generated more task-relevant and fewer task-irrelevant options than their less-skilled counterparts. Likewise, the numbers of task-relevant and task-irrelevant options generated were more predictive of anticipation accuracy and selection of the criterion best option. This finding offers support for the use of LTWM-based mechanisms to facilitate skilled option generation, at least for the representative soccer tasks used in this line of research.

However, a few interesting findings offered some reconciliation between LTWM and TTF. During intervention trials (but not assessment trials) and in line with TTF, skilled participants generated fewer options in total than their less-skilled counterparts (see Raab & Johnson, 2007) and performance trended negatively—albeit non-significantly—with the total number of options generated (see Johnson & Raab, 2003). Furthermore, a selective reduction in the type of information generated was only observed during intervention—such that additional time permitted the generation of additional irrelevant (as opposed to relevant) information. This unique finding was also observed in Experiment 2. Taken together, these results suggested that although the LTWM-based claims of Ward and colleagues (2013) offers the better *explanation* of skilled option generation during assessment and intervention, the *prescriptions* of TTF may still hold during intervention-phase trials. In other words, these results suggest that the participants would have benefitted from acting on an initial intuitive response during

intervention, as opposed to deliberating between multiple, task-relevant or otherwise, options that were recognized in the environment/situation. Nevertheless, the ability to generate more (as opposed to fewer) task-relevant options was consistently related to higher performance and skill.

Experiments 4 and 5 also provided theoretical contribution. The Online Assessment of Strategic Skill In Soccer (OASSIS) was designed based off of previous research on anticipation and temporal occlusion in sport (e.g., Abernethy 1990; Abernethy & Russell, 1987; Ward et al., 2013). While this tool was developed for application beyond research purposes, it also offered an independent assessment of perceptual-cognitive skill with which I could compare option generation behavior. Again, support for the LTWM-based claims of Ward and colleagues (2013) was provided. Participants' performance on the OASSIS was significantly and positively related to the number of task-relevant options and trended negatively with the number of task-irrelevant options that participants generated on the option generation task. This was at least the case for Experiment 5, which utilized highly skilled participants. Performance on the OASSIS was not negatively related to the total number of options generated on the option generation paradigm (cf. Johnson & Raab, 2003) during Experiment 4. When taken together, these findings—in addition to those of Experiments 1, 2, and 3 provide support for the notion that performance in these types of domains is facilitated by LTWM-type mechanisms (as opposed to a simpler more intuitive heuristic such as TTF).

In addition to testing the claims of extant theory, this body of work was committed to the development of an applied tool for assessing perceptual-cognitive skill in soccer. The OASSIS was created using trials that employed a similar temporal occlusion task to the option generation paradigm, but adhered to a multiple choice format. In Experiments 4 and 5, I demonstrated that the OASSIS can predict skill (i.e., known-groups validity) and performance (i.e., predictive validity) on the option generation paradigm. This was shown in high- and low-skill groups. Moreover, the relationships between performance on the OASSIS and option generation behavior demonstrate convergent validity for the OASSIS as a test of perceptual-cognitive skill.

In Experiment 6, I demonstrated that the OASSIS accounts for more variance between skill groups than measures of domain-general cognition. This included the BNT, which was demonstrated to be related to the cognitive attributes highlighted by Voss et al. (2010), and the MRT-A, which has been validated as a measure of domain-general spatial skill. This is in line with previous research on the relative contribution of domain-specific skill and domain-general abilities (e.g., Helsen & Starkes, 1999; Ward & Williams, 2003). However, I also sought to compare performance on OASSIS with domain-specific knowledge (as assessed by the soccer knowledge questionnaire). The relationship between the two demonstrated further convergent validity for the OASSIS as a perceptual-cognitive assessment tool in soccer, but also revealed a troublesome finding for the OASSIS. The soccer knowledge questionnaire accounted for more variance than OASSIS performance, even to the extent that OASSIS was no longer a significant

predictor in the regression model. I suspect that this is the result of comparing recreational level and Division I soccer players without an intermediate skill group. While knowledge of the game and rules may separate expert players from novices, I suspect it would not separate expert players from near-experts. For this reason, I advise that future research compare the relative contribution of the OASSIS and domain-specific knowledge to predicting membership among three or more skill groups.

Furthermore, future research should investigate the specific items of the OASSIS. Given the wide variety of stimuli, it is very likely that certain trials are more useful than others for predicting skill. By organizing the trials based on an adaptive structure—where participants are guided to the most predictive trials based on their performance thus far—the predictive validity of OASSIS in regard to skill could be increased. Furthermore, the utility of the soccer knowledge questionnaire in discriminating between skill groups suggest that perhaps knowledge-related questions could be included as part of a more comprehensive assessment of cognition in soccer.

This body of work has provided an avenue for future research. On one hand, the contribution to theoretical accounts of skilled option generation has implications for training that must be empirically evaluated. During assessment-phase tasks, training programs should focus on the generation of as many threatening (i.e., task-relevant) alternatives as are available in the environmental/situational structure, in addition to the actual situational outcome, all while reducing the generation of irrelevant information. Such a training program would inform the mental situational-model-building

process described by LTWM theory (see Ericsson & Kintsch, 1995; Ward et al., 2013).

During intervention-phase tasks, training should emphasize the generation of a quick and intuitive response in line with TTF (see Johnson & Raab, 2003) that contains high-quality (i.e., task-relevant) options, to the extent that they are available in the environment.

However, the results of this body of research suggest that the generation of only irrelevant information should be reduced—as opposed to reducing all option generation down to a single option (cf. Johnson & Raab, 2003).

On one hand, these implications for theory and training offer substantial contribution to the state-of-the-science. Instead of the proponents of LTWM and TTF continuing to work in relative isolation (at least from a methodological standpoint), this research has defined boundary conditions for these separate cognitive mechanisms (e.g., assessment and intervention). The task which is now presented to researchers in naturalistic decision making is to test the claims of these models across a number of different domains/situations of interest, as I have done in these representative soccer tasks. I speculate that these environmental constraints (e.g., decision perspective and time pressure) may have similar effects on option generation behavior in law enforcement, military, and many other complex and dynamic domains. Likewise, I speculate that the situational-model-building process described by LTWM (see Ericsson & Kintsch, 1995; Ward et al., 2013) may provide the best explanation of skilled option generation, anticipation, and decision making in other such environments. Similarly, I speculate that

the prescriptions of TTF may hold—at least during the intervention phase—across other similar domains. Of course further work is certainly needed to substantiate this claim.

On the other hand, this body of work has provided an avenue for research in applied tools. Outside of the suggestions for future work on the OASSIS, similar tools—based on temporal occlusion during representative tasks—can be designed in a number of domains. Such tools for assessing perceptual-cognitive skill in complex and dynamic domains can also be used to evaluate the effectiveness of aforementioned training programs based on option generation behavior. Ultimately, this body of work offers structure with which to move forward when assessing and training skilled decision making in a number of real-world domains. Training and individual assessments based on option generation behavior, anticipation, and decision making may accelerate the onset of expertise across of a number of these domains—thereby increasing the success and efficiency of those who operate within them and reducing the prevalence of costly errors—whether measured in time, resources, or even lives—such as those committed by the crew of the USS Vincennes.



## References

- Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sport Sciences*, 8(1), 17-34.
- Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science*, 6(4), 283-319.
- Allard, F., Graham, S., & Paarsalu, M. E. (1980). Perception in sport: Basketball. *Journal of Sport Psychology*, 2(1), 14-21.
- Allard, F., & Starkes, J. L. (1980). Perception in sport: Volleyball. *Journal of Sport Psychology*, 2(1), 22-33.
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89, 369-406.
- Belling, P. K., Suss, J., & Ward, P. (under review A). Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making and option generation behavior. *Psychology of Sport and Exercise*.
- Belling, P. K., Suss, J., & Ward, P. (under review B). The effect of time constraint on anticipation, decision-making and option generation in complex and dynamic environments. *Cognition, Technology and Work*.
- Belling, P. K., Suss, J. & Ward, P. (2013). Investigating constraints on decision making strategies. In Chaudet, H., Pellegrin, L., & Bonnardel, N. (Eds.) *Proceedings of the 11<sup>th</sup> International Conference on Naturalistic Decision Making (NDM 2013)*,

- Marseille, France, 21-24 May 2013.* (pp. 133-136). Paris, France: Arpege Science Publishing.
- Belling, P. K., & Ward, P. (2012). Evaluating the Take-The-First heuristic in assessing situations and decision making using an option generation paradigm in soccer. *Journal of Sport & Exercise Psychology, 34*, S210.
- Belling, P. K., & Ward, P. (in press). The effectiveness of an online test of anticipation at predicting skill and option generation behavior. Oral Presentation at *NASPSPA 2013 Annual Meeting*.
- Borgeaud, P., & Abernethy, B. (1987). Skilled perception in volleyball defense. *Journal of Sport Psychology, 9*(4), 400-406.
- Burroughs, W. A. (1984). Visual simulation training of baseball batters. *International Journal of Sport Psychology, 15*, 117-126.
- Calderwood, R., Klein, G. A., & Crandall, B. W. (1988). Time pressure, skill, and move quality in chess. *The American Journal of Psychology, 101*(4), 481-493.
- Chabris, C. F., & Hearst, E. S. (2003). Visualization, pattern recognition, and forward search: Effects of playing speed and sight of the position on grandmaster chess errors. *Cognitive Science, 27*(4), 637-648.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.
- Cokely, E.T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin Numeracy Test. *Judgment and Decision Making, 7*, 25-47.

- Cox, R. H., Russell, W. D., & Robb, M. (1999). Comparative concurrent validity of the MRF-L and ARS competitive state anxiety rating scales for volleyball and basketball. *Journal of Sport Behavior*, 22(3), 310-320.
- Ericsson, K.A., & Kintsch, W. (1995). Long Term Working Memory. *Psychological Review* 102(2), 211-245.
- Ericsson, K. A., Krampe, R. T., Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Ericsson, K. A., & Ward, P. (2007). Capturing the Naturally Occurring Superior Performance of Experts in the Laboratory. *Association for Psychological Science*, 16(6), 346-350.
- Fadde, P. J. (2006). Interactive video training of perceptual decision making in the sport of baseball. *Technology, Instruction, Cognition and Learning*, 4(3), 265-285.
- Farrow, D., Abernethy, B., Jackson, R. C. (2005). Probing expert anticipation with the temporal occlusion paradigm: Experimental investigations of some methodological issues. *Motor Control*, 9(3), 330-349.
- French, K. E., & Thomas, J. R. (1987). The relation of knowledge development to children's basketball performance. *Journal of Sport Psychology*, 9(1), 15-32.
- Gabbett, T., Rubinoff, M., Thorburn, L., & Farrow, D. (2009). Testing and training anticipation skills in softball fielders. *International Journal of Sports Science and Coaching*, 2(1), 15-24. DOI: 10.1260/174795407780367159.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox.

- In G. Gigerenzer, P. M. Todd, and the ABC Research Group. *Simple heuristics that make us smart* (pp. 3-34). New York: Oxford University Press.
- Gobet, F., & Simon, H. A. (1996a). The roles of recognition processes and look-ahead search in time constrained expert problem solving: Evidence from grand-master-level chess. *Psychological Science* 7(1).
- Gobet, F. & Simon, H. A. (1996b). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, 31, 1-40.
- Gorman, A. D., Abernethy, B., & Farrow, D. (2013). Is the relationship between pattern recall and decision making influenced by anticipatory recall? *The Quarterly Journal of Experimental Psychology*, (in-press), 1-42.
- Goulet, C., Bard, C., & Fleury, M. (1989). Expertise differences in preparing to return a tennis serve: a visual information processing approach. *Journal of Sport & Exercise Psychology*, 11(4), 382-398.
- Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task induction. *Journal of Experimental Psychology: Applied*, 5(2), 129-151.
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: an investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44(4), 339-387.
- Haskins, M. J. (1965). Development of a response-recognition training film in tennis. *Perceptual and Motor Skills*, 21, 207-211.

- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press, Inc.
- Helsen, W., & Pauwels, J. M. (1988) The use of a simulator in evaluation and training of tactical skills in football. In T. Reilly, A. Lees, K. Davids, & W. J. Murphy (Eds.), *Science and football* (pp. 493-497). London: E & FN Spon.
- Helsen, W. F., & Starkes, J. L. (1999). A multidimensional approach to skilled perception and performance in sport. *Applied Cognitive Psychology*, 13(1), 1-27.
- Howell, D. C. (2002). *Statistical Methods for Psychology* (5<sup>th</sup> ed.). Pacific Grove, CA: Duxberry.
- Johnson, J. G., & Raab, M. (2003). Take The First: Option-generation and Resulting Choices. *Organizational Behavior and Human Decision Processes* 91, 215-29.
- Jones, C. M., & Miles, T. R. (1978). Use of advance cues in predicting the flight of a lawn tennis ball. *Journal of Human Movement Studies*, 4(4), 231-235.
- Joslyn, S., & Hunt, E. (1998). Evaluating individual differences in response to time-pressure situations. *Journal of Experimental Psychology: Applied*, 4(1), 16-43.
- Klein, G. A. (1989). *Recognition-primed decisions*. Defense Technical Information Center.
- Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsombok (Eds.), *Decision making in action: Models and methods* (pp.138–147). Norwood, NJ: Ablex.

- Klein, G.A. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes*, 62(1), 63-69.
- Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *Systems, Man and Cybernetics, IEEE Transactions on*, 19(3), 462-472.
- Klein, G. A., Orasanu, J., Calderwood, R., & Zsombok, C. E. (1993). *Decision making action: Models and methods* (pp. 138-147). Norwood, NJ: Ablex.
- Klein, G. A., & Peio, K. J. (1989). Use of a prediction paradigm to evaluate proficient decision making. *The American journal of psychology*, 321-331.
- Lassiter, G. D. (2000). The relative contributions of recognition and search-evaluation processes to high-level chess performance: Comment on Gobet and Simon. *Psychological Science*, 11(2), 172-173.
- Mann, D. L., Farrow, D., Shuttleworth, R., Hopwood, M., & MacMahon, C. (2009). The influence of viewing perspective on decision making and visual search behaviour in an invasive sport. In *Skill Acquisition Research and Application in Australasia. Inaugural Conference of the Australasian Skill Acquisition Research Group (ASARG), Melbourne, Australia, 15-16 June 2007*. (Vol. 40, No. 4, pp. 546-564). Edizioni Luigi Pozzi.
- Mann, D. T. Y., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, 29(4), 457.

- McPherson, S. L. & French, K.E. (1991). Changes in cognitive strategies and motor skill in tennis, *Journal of Sport and Exercise Psychology*, 13(1), 26-41.
- McPherson, S. L., & Thomas, J. R. (1989). Relation of knowledge and performance in boys' tennis: Age and expertise. *Journal of Experimental Child Psychology*, 48(2), 190-211.
- North, J. S., Ward, P., Ericsson, A., & Williams, A. M. (2011). Mechanisms underlying skilled anticipation and recognition in a dynamic and temporally constrained domain. *Memory*, 19(2), 155-168.
- North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied cognitive psychology*, 23(6), 878-894.
- Osman, A., Lou, L., Muller-Gethman, H., Rinkenauer, G., Mattes, S., & Ulrich, R. (2000). Mechanisms of speed-accuracy tradeoff: Evidence from covert motor processes. *Biological Psychology*, 51, 173-199.
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition* 28, 39-58.
- Proctor, R. W., & Van Zandt, T. (2008). *Human factors in simple and complex systems*. CRC Press Llc.
- Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-

- generation strategies. *Journal of Experimental Psychology: Applied* 13(3), 158-170.
- Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1–48.
- Ripoll, H., Kerlirzin, Y., Stein, J. F., & Reine, B. (1995). Analysis of information process, decision making, and visual strategies in complex problem solving sport situations. *Human Movement Science*, 14(3), 325-349.
- Rosalie, S. M., & Müller, S. (2013). Timing of in-situ visual information pick-up that differentiates expert and near-expert anticipation in a complex motor skill. *The Quarterly Journal of Experimental Psychology*, DOI: 10. 1080/ 17470218. 2013. 770044.
- Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: recent developments in quantitative methods for literature reviews. *Annual Review of Psychology*, 52, 59-82.
- Savelsbergh, G. J. P., Williams, A. M., Van Der Kamp, J., & Ward, P. (2002). Visual search, anticipation, and expertise in soccer goalkeepers. *Journal of Sports Sciences*, 20, 279-287.
- Shepard, R & Metzler, J. (1971). Mental rotation of three dimensional objects. *Science* 171(972), 701-703.
- Simon, H.A. (1969). *The Sciences of the Artificial*. Cambridge, MA: MIT Press
- Starkes, J. L. (1987). Skill in field hockey: The nature of the cognitive advantage.



*Journal of Sport Psychology*, 9(2), 146-160.

- Starkes, J. L., & Deakin, J. M. (1984). Perception in sport: A cognitive approach to skilled performance. In Straub, W. F., & Williams, J. M. (Eds.), *Cognitive Sport Psychology*. Lansing, NY: Sports Science Associates.
- Starkes, J. L., Edwards, P., Dissanayake, P., & Dunn, T. (1995). A new technology and field test of advance cue usage in volleyball. *Research quarterly for exercise and sport*, 66(2), 162.
- Tenenbaum, G., Yuval, R., Elbaz, G., Bar-Eli, M., & Weinberg, R. (1993). The relationship between cognitive characteristics and decision making. *Canadian Journal of Applied Physiology*, 18(1), 48-62.
- Travassos, B., Araujo, D., Davids, K., O'Hara, K., Leitao, J., Cortinhas, A. (2013). Expertise effects on decision making in sport are constrained by requisite response behaviours – A meta-analysis. *Psychology of Sport and Exercise*, 14(2), 211-219.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Unsworth, N., & Spillers, G.J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, 62, 392–406.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 69, 915-921.

- von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.
- Voss, M. W., Kramer, A. F., Basak, C., Prakash, R. S., & Roberts, B. (2010). Are expert athletes 'expert' in the cognitive laboratory? A meta-analytic review of cognition and sport expertise. *Applied Cognitive Psychology*, 24, 812-826. DOI: 10.1002/acp.1588.
- Ward, P. (2002). The development of perceptual-cognitive expertise. Unpublished doctoral thesis, Liverpool John Moores University, UK.
- Ward, P., Ericsson, K.A., & Williams, A.M. (2013). Complex perceptual-cognitive expertise in a simulated task environment. *Journal of Cognitive Engineering and Decision Making*, 7, 231-254.
- 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(1), S71-S102. DOI:10.1080/08995600701804814.
- Ward, P., Gabbard, S. R., Watamaniuk, S. N. J., Macquet, A. C., Araujo, D., Veinott, E. S., Mueller, S. T., & Fadde, P. (2011) Symposium on recognition, prediction and decision making in sport. In *Proceedings of the 10<sup>th</sup> International Conference on Naturalistic Decision Making*, May 31-June 3, 2011. Orlando, FL.
- Ward, P., Suss, J., Eccles, D. W., Williams, A. M., & Harris, K. R. (2011). Skill-based

- differences in option generation in a complex task: A verbal protocol analysis. *Cognitive Processing: International Quarterly of Cognitive Science*, 12, 289-300.
- Ward, P., & Williams, A. M. (2003). Perceptual and cognitive skill development in soccer: The multidimensional nature of expert performance. *Journal of Sport & Exercise Psychology*, 25, 93-111.
- Williams, A. M. (2000). Perceptual skill in soccer: Implications for talent identification and development. *Journal of Sports Sciences*, 18(9), 737-750.
- Williams, A. M., & Davids, K. (1995). Declarative knowledge in sport: A byproduct of experience or a characteristic of expertise? *Journal of Sport & Exercise Psychology*, 17(3), 259-275.
- Williams, A. M., Davids, K., Burwitz, L., Williams, J. (1993). Cognitive knowledge and soccer performance. *Perceptual and Motor Skills*, 76, 579-593.
- Williams, A. M., & Ward, P. (2007). Anticipation and decision making: Exploring new horizons. *Handbook of Sport Psychology, Third Edition*, 203-223.
- Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied*, 8(4), 259.
- Wright, D. L., Pleasants, F., & Gomez-Meza, M. (1990). Use of advanced visual cue sources in volleyball. *Journal of Sport & Exercise Psychology* 12(4), 406-414.
- Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008).

Training generalized spatial skills. *Psychonomic Bulletin & Review*, 15(4), 763-771.

Zijlstra, F. R. H. (1993). Efficiency in work behaviour: A design approach for modern tools (Ph.D. Thesis).

Zsombok, C. E., & Klein, G. (1997). *Naturalistic Decision Making*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

## **Appendix**

### **Supplemental Figures and Tables**

Participant \_\_\_\_\_

**Soccer experience questionnaire**

- 1) What is your gender?                    *M*        /        *F*
- 2) What is your date of birth?        \_\_\_\_/\_\_\_\_/\_\_\_\_
- 3) How old were you when you first started playing soccer for fun?
- 4) How old were you when you started playing soccer seriously (i.e., received coaching and joined a team)?
- 5) Most often, what position do you play? If you play two positions equally, which do you prefer?
  - a. Defense
  - b. Midfield
  - c. Forward
  - d. Goalkeeper
- 6) What is the highest level of competition at which you have played soccer (e.g., high school varsity, college varsity)?
  - a. Have not played organized soccer
  - b. Recreational or intramural level
  - c. High school varsity level
  - d. College club level
  - e. College varsity level
  - f. Other. Please specify \_\_\_\_\_
- 7) On average, how many **organized** (i.e., with a club team) competitive matches do you play
  - a. During this year?
  - b. During the last three years?
  - c. Prior to the last three years?
- 8) On average, how many **non-organized** (i.e., pick-up) matches do you play
  - a. During this year?
  - b. During the last three years?
  - c. Prior to the last three years?

**Figure 7.** The Soccer Experience Questionnaire (p. 1).

9) On average, how many hours per week did you spend practicing (not playing) in a **non-team setting**

- a. During this year?
- b. During the last three years?
- c. Prior to the last three years?

10) On average, how many hours per week did you spend practicing in a **team setting** (i.e., with a club)

- a. During this year?
- b. During the last three years?
- c. Prior to the last three years?

11) On average, how many competitive matches do you **watch**, either in person, or on television?

- a. During this year?
- b. During the last three years?
- c. Prior to the last three years?

12) Have you played other sports besides soccer? If so, at what level of competition?

**Figure 8.** The Soccer Experience Questionnaire (p. 2).

**Table 13.** Means (SD) frequency scores for all performance measures, main effects of skill, and interaction effects of skill by time during six time constrained and six non-time constrained assessment trials of Experiment 1.

	Anticipation accuracy of actual outcome		Generation of criterion most threatening option		Accurate rating of criterion most threatening option		Anticipation of criterion most threatening option	
	TC	NTC	TC	NTC	TC	NTC	TC	NTC
Low-skill	0.85 (0.55)	0.85 (0.69)	1.92 (1.44)	2.53 (1.44)	1.54 (1.45)	2.08 (1.04)	1.38 (1.45)	1.77 (1.30)
Moderate-skill	1.39 (0.99)	1.39 (0.78)	2.87 (1.36)	3.48 (1.34)	2.43 (1.34)	2.57 (1.31)	1.96 (1.36)	2.30 (1.26)
Skill effect	$F(1, 34) = 17.97$ $p < 0.01$ $\eta p^2 = 0.35$		$F(1, 34) = 4.21$ $p = 0.01$ $\eta p^2 = 0.17$		$F(1, 34) = 3.78$ $p = 0.06$ $\eta p^2 = 0.06$		$F(1, 34) = 2.31$ $p = 0.14$ $\eta p^2 = 0.06$	
Time effect	$F(1, 34) = 0.00$ $p = 1.00$ $\eta p^2 < 0.01$		$F(1, 34) = 4.21$ $p = 0.05$ $\eta p^2 = 0.11$		$F(1, 34) = 1.44$ $p = 0.24$ $\eta p^2 = 0.04$		$F(1, 34) = 1.64$ $p = 0.21$ $\eta p^2 = 0.05$	
Skill x Time interaction effect	$F(1, 34) = 0.00$ $p = 1.00$ $\eta p^2 < 0.01$		$F(1, 34) = 0.00$ $p = 0.99$ $\eta p^2 < 0.01$		$F(1, 34) = 0.54$ $p = 0.47$ $\eta p^2 = 0.02$		$F(1, 34) < 0.01$ $p = 0.95$ $\eta p^2 < 0.01$	

**Table 14.** Means (SD) frequency scores for all performance measures, main effects of skill, and interaction effects of skill by time during six time constrained (TC) and six non-time constrained (NTC) intervention trials of Experiment 1.

	Generation of criterion best option		Accurate rating of criterion best option		Selection of criterion best option	
	Time constraint	No time constraint	Time constraint	No time constraint	Time constraint	No time constraint
Low-skill	1.85 (1.14)	1.77 (1.09)	1.15 (0.90)	1.23 (0.72)	1.31 (1.11)	1.46 (0.97)
Moderate-skill	2.52 (1.31)	2.43 (1.08)	1.78 (1.17)	1.48 (0.99)	1.87 (1.28)	1.56 (1.20)
Skill effect	$F(1, 34) = 6.19$ $p < 0.01, \eta p^2 = 0.15$		$F(1, 34) = 3.64$ $p = 0.07, \eta p^2 = 0.07$		$F(1, 34) = 1.65$ $p = 0.21, \eta p^2 = 0.05$	
Time effect	$F(1, 34) = 0.07$ $p = 0.79, \eta p^2 < 0.01$		$F(1, 34) = 0.19$ $p = 0.66, \eta p^2 = 0.06$		$F(1, 34) = 0.06$ $p = 0.81, \eta p^2 < 0.01$	
Skill x Time interaction effect	$F(1, 34) = 0.00$ $p = 0.99, \eta p^2 < 0.01$		$F(1, 34) = 0.54$ $p = 0.48, \eta p^2 = 0.02$		$F(1, 34) = 0.53$ $p = 0.47, \eta p^2 = 0.02$	



**Table 15.** Mean (SD) frequency scores of all performance measures during six time constrained and six non-time constrained assessment trials and main effects of time constraint during Experiment 2.

	Anticipation accuracy of actual outcome	Generation of most threatening option	Accurate rating of most threatening option	Anticipation of most threatening option
Time constraint	1.33 (1.28)	2.05 (1.72)	1.62 (1.36)	1.57 (1.29)
No time constraint	1.43 (1.12)	1.95 (1.63)	1.86 (1.62)	1.62 (1.47)
Effect of time constraint	$t(1, 20) = 0.22$ $p = 0.83$ $d = 0.08$	$t(1, 20) = 0.22$ $p = 0.83$ $d = 0.06$	$t(1, 20) = -0.57$ $p = 0.58$ $d = -0.16$	$t(1, 20) = -0.12$ $p = 0.91$ $d = -0.03$

**Table 16.** Mean (SD) frequency scores of all performance measures during six time constrained and six non-time constrained intervention trials and main effects of time constraint during Experiment 2.

	Generation of criterion best option	Accurate rating of criterion best option	Selection of criterion best option
Time Constraint	1.76 (1.18)	1.14 (1.15)	1.33 (1.15)
No Time Constraint	1.48 (1.03)	0.81 (0.87)	0.90 (1.14)
Effect of Time Constraint	$t(1, 20) = 0.74$ $p = 0.47, d = 0.25$	$t(1, 20) = 1.10$ $p = 0.29, d = 0.32$	$t(1, 20) = 1.34$ $p = 0.20, d = 0.38$

**Table 17.** Correlation coefficients between option generation behavior and performance measures during time constrained assessment trials during Experiment 2.

	Total options	Task-relevant options	Task-irrelevant options
Generation of criterion most threatening option	$r = 0.22$ $p = 0.32$	$r = 0.36$ $p = 0.11$	$r < 0.01$ $p = 0.99$
Accurate rating of criterion most threatening option	$r = 0.02$ $p = 0.94$	$r = 0.20$ $p = 0.39$	$r = -0.13$ $p = 0.59$
Anticipation of criterion most threatening option	$r = 0.11$ $p = 0.65$	$r = 0.18$ $p = 0.43$	$r = -0.01$ $p = 0.98$
Anticipation accuracy of actual outcome	$r = 0.35$ $p = 0.13$	$r = 0.40$ $p = 0.07$	$r = 0.12$ $p = 0.60$

**Table 18.** Correlation coefficients between option generation behavior and performance measures during non-time constrained assessment trials during Experiment 2.

	Total options	Task-relevant options	Task-irrelevant options
Generation of criterion most threatening option	$r = -0.09$ $p = 0.70$	$r = 0.60$ $p < 0.01$	$r = -0.48$ $p = 0.03$
Accurate rating of criterion most threatening option	$r = -0.11$ $p = 0.62$	$r = 0.59$ $p < 0.01$	$r = -0.49$ $p = 0.02$
Anticipation of criterion most threatening option	$r = -0.15$ $p = 0.53$	$r = 0.45$ $p = 0.04$	$r = -0.43$ $p = 0.05$
Anticipation accuracy of actual outcome	$r = 0.09$ $p = 0.71$	$r = 0.30$ $p = 0.19$	$r = -0.11$ $p = 0.64$

**Table 19.** Correlation coefficients between option generation behavior and performance measures during time constrained intervention trials during Experiment 2.

	Total options	Task-relevant options	Task-irrelevant options
Generation of criterion best option	$r = 0.50$ $p = 0.02$	$r = 0.72$ $p < 0.01$	$r = -0.09$ $p = 0.69$
Accurate rating of criterion best option	$r = 0.11$ $p = 0.65$	$r = 0.50$ $p = 0.02$	$r = -0.39$ $p = 0.08$
Selection of criterion	$r = 0.28$	$r = 0.63$	$r = -0.27$

best option	$p = 0.21$	$p < 0.01$	$p = 0.23$
-------------	------------	------------	------------

**Table 20.** Correlation coefficients between option generation behavior and performance measures during non-time constrained intervention trials during Experiment 2.

	Total options	Task-relevant options	Task-irrelevant options
Generation of criterion	$r = 0.11$	$r = 0.36$	$r = -0.23$
best option	$p = 0.63$	$p = 0.11$	$p = 0.32$
Accurate rating of	$r = -0.36$	$r = -0.17$	$r = -0.37$
criterion best option	$p = 0.11$	$p = 0.46$	$p = 0.10$
Selection of criterion	$r = -0.20$	$r < 0.01$	$r = -0.33$
best option	$p = 0.38$	$p = 0.99$	$p = 0.15$

- 1) In soccer, what part of the body is not allowed to touch the ball?
  - a. Head
  - b. Arm
  - c. Leg
  - d. Chest
- 2) Please select each of the statements below that are TRUE regarding a legal throw in during a soccer match. Note there may be more than one TRUE statement.
  - a. The ball must go completely behind the thrower's head during the throwing motion.
  - b. The ball must be out of bounds for 5 seconds before it can be thrown back in.
  - c. Both feet must be touching the ground during the throwing motion.
  - d. Only one hand is allowed to be touching the ball during the throwing motion.
  - e. Both hands must be touching the ball during the throwing motion.
  - f. The thrower is allowed to throw the ball to himself.
- 3) In a standard adult-level soccer game, how many players are on the field (including goalkeeper) for each team?
  - a. 7
  - b. 9
  - c. 11
  - d. 13
  - e. 15
  - f. 17
- 4) When a player is kicking a free kick, what is the minimum distance allowed between opposing team members and the ball?
  - a. 20 yards
  - b. 15 yards
  - c. 10 yards
  - d. 5 yards
  - e. There is no minimum distance.
- 5) The penalty spot is a marking on the field where penalty kicks are taken from. How far is the penalty spot from the goal?
  - a. 10 yards
  - b. 12 yards
  - c. 14 yards
  - d. 16 yards
  - e. 18 yards
  - f. 20 yards

**Figure 9.** Soccer Knowledge Questionnaire (p. 1).

- 6) When a team kicks the ball out of bounce behind the OTHER team's goal, the resulting action is a \_\_\_\_\_.
- a. Goal kick
  - b. Corner kick
  - c. Penalty kick
  - d. Drop ball
- 7) How long do professional soccer games last (excluding extra time)?
- a. 45 minutes
  - b. 60 minutes
  - c. 90 minutes
  - d. 120 minutes
- 8) In competitive soccer leagues, players are typically required to wear? (Select all that apply)
- a. Shoes with legal cleats
  - b. Shin guards
  - c. Thigh pads
  - d. Shoulder pads
  - e. A helmet
  - f. Protective gloves

**Figure 10.** Soccer Knowledge Questionnaire (p. 2).