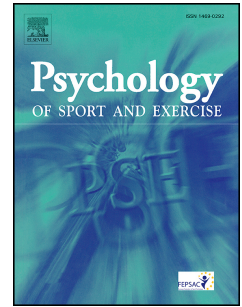


Accepted Manuscript

Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option-generation behavior

Patrick K. Belling, Joel Suss, Paul Ward



PII: S1469-0292(14)00108-3

DOI: [10.1016/j.psychsport.2014.08.001](https://doi.org/10.1016/j.psychsport.2014.08.001)

Reference: PSYSPO 920

To appear in: *Psychology of Sport & Exercise*

Received Date: 1 October 2013

Revised Date: 21 July 2014

Accepted Date: 1 August 2014

Please cite this article as: Belling, P.K., Suss, J., Ward, P., 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 & Exercise* (2014), doi: 10.1016/j.psychsport.2014.08.001.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option-generation behavior

Patrick K. Belling^a, Joel Suss^b, and Paul Ward^c

^aMichigan Technological University, USA

^bUniversité Laval, Canada

^cUniversity of Greenwich, UK

Author Note and Acknowledgements

Patrick Belling and Joel Suss are in the Department of Cognitive and Learning Sciences at Michigan Technological University. Paul Ward is Head of Centre for Sports Science and Human Performance at University of Greenwich, Kent, UK.

The results presented in this manuscript contain extended analyses and report of data presented at the 11th international conference on Naturalistic Decision Making in Marseille, France (see Belling, Suss, & Ward, 2013).

Correspondence should be directed to Patrick Belling, Department of Cognitive and Learning Sciences, Michigan Technological University, 1400 Townsend Dr., Houghton, MI 49931-1295. Tel: 906-487-2460. Fax: 906-487-2468. Email: pkbellin@mtu.edu.

We would also like to acknowledge the U.S. Soccer Federation and the NASPSPA Graduate Student Research Grant for providing funding which supported parts of the research presented in this manuscript.

Abstract

Objectives. Three main goals were addressed in this research. First, we tested the claims of two cognitive mechanisms that have been proposed to explain expert performance. This was done during assessment and intervention phases of decision making. Second, we tested the validity of an online test of perceptual-cognitive skill in soccer: The Online Assessment of Strategic Skill In Soccer (OASSIS). Third, we compared the OASSIS to other predictors of skill in soccer. Design. Over the course of a three-part experiment, participants completed an updated version of the option-generation paradigm employed by Ward, Ericsson, and Williams (2013), the OASSIS, and a battery of other cognitive tests. Performance on these tests was used to inform theory and validate the OASSIS as an applied tool for domain professionals. Method. NCAA Division 1 and recreational-level soccer players completed a battery of tests, both using paper/pencil (see Ward et al., 2013) and online. Results. Support for Long Term Working Memory theory (LTWM; see Ericsson & Kintsch, 1995) was observed during both phases of decision making, though the prescriptions of the Take-The-First heuristic (see Johnson & Raab, 2003) tend to hold, particularly within intervention phase. When used to predict skill-group membership, the OASSIS accounted for more variance than other domain-general tests of cognition. Furthermore, scores on the OASSIS correlated with other measures of perceptual-cognitive skill in soccer and the process-level predictions made by LTWM. Conclusions. Updates to our theoretical understanding of expert performance are provided and the validity of the OASSIS is demonstrated.

Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option-generation behavior

A large body of research has demonstrated the effectiveness of anticipation (e.g., Williams & Davids, 1995) and decision-making ability (Gorman, Abernethy & Farrow, 2013; Raab & Johnson, 2007) as predictors of skill level in sport. However, little, if any, research has examined both anticipation and decision-making skill within the same player. This is troublesome considering that recent work has conceptualized these two skills as critical subcomponents of the decision process in naturalistic and/or complex domains (Belling & Ward, 2012; Ward, Suss, Eccles, Williams, & Harris, 2011), and more generally as reciprocal phases of the perception-action cycle (Engström, Kelso, & Holroyd, 1996; Neisser, 1976). Ward et al. (2011) conceptualize the decision process as two co-occurring phases: an assessment phase—where decision makers assess the situation, generate options, and anticipate the course(s) of action to be taken by others in the environment; and an intervention phase—where they generate options and select the course(s) of action to pursue themselves. Typically, these phases of decision making have been investigated in isolation and different theoretical accounts have been offered to explain superior performance in each phase (e.g., Raab & Johnson, 2007; Ward, Ericsson, & Williams, 2013). Consequently, our first goal was to examine whether similar or dissimilar mechanisms support superior performance in each decision making phase to better understand the strategies employed by skilled decision makers in the types of sporting environments examined, and the training needs of those who are less skilled. In this particular research, the sport investigated is soccer, though the theoretical and practical claims are likely to extend to other sports and/or domains that have similar characteristics to sport (e.g., high speed decision-making under uncertainty).

An updated option generation task is employed in this research. This is similar to the option generation tests used in previous sport research (e.g., Johnson & Raab, 2003; Ward et al., 2013), where decision-makers generate situational sport outcomes as part of a representative task. This task is described in more detail below. Unfortunately, the complexity and time-consuming nature of current, albeit novel, methods used to elucidate the cognitive strategies supporting superior performance, while enlightening, are unlikely to be used for diagnostic purposes (i.e., to assess anticipation and decision-making skill and to identify deficiencies in strategies used). This reduces their efficacy in terms of tailoring perceptual-cognitive training to the specific needs of individuals. Our second goal was, therefore, to develop a simpler test of perceptual-cognitive skill capable of discriminating between skill groups. Our aim was to assess the predictive validity of the simpler test in terms of performance on the more

complex option generation test, and examine construct validity of the simple test by demonstrating its relationship with option generation strategy use. A tool of this kind, however, would only be useful if it predicted skill level over and above other potential predictors relevant to soccer. Hence, a third and final goal was to examine the relative contribution of the simple perceptual-cognitive skill test and domain-general cognitive abilities to soccer skill.

In this paper, we conducted a three-part experiment to empirically examine each of these goals. In Experiment 1A, we tested current models of skilled decision making to delineate the mechanisms used to explain superior performance in the assessment and intervention phases of decision making (as conceptualized by Ward et al., 2011) in the specific types of situations tested. In Experiment 1B, we developed a shorter, simpler, online tool (online assessment of strategic skill in soccer; OASSIS) to assess perceptual-cognitive skill in soccer. We examined its predictive validity (relative to the test of option generation) and construct validity, specifically its convergence with task-relevant option generation and divergence with task-irrelevant option generation. In Experiment 1C, we further examined whether the short, online tool was a better predictor of skill level than other domain-general cognitive predictors. We begin by examining the research on perceptual-cognitive skill that has investigated the assessment and intervention phases of decision making.

EXPERIMENT 1A

Tests of Decision Making in Sport

Assessment phase decision making. There has been an emphasis in the sports expertise literature over the last half decade on the perceptual and memory-related aspects of the assessment phase of decision making. In particular, there has been a strong focus on identifying skill-based differences in anticipation performance and improving players' anticipation skill (for a review, see Ward, Williams, & Hancock, 2006). Since Haskins' (1965) paper investigating anticipation in tennis, researchers have used experimental methods, particularly temporal occlusion¹, to investigate how players anticipate what their opponent will do next. For instance, Jones and Miles

¹ In the traditional temporal-occlusion approach, participants watch video clips that depict typical game play within that sport. In the videos, information is presented up until a critical decision point at which time/point the presentation of additional information (i.e., more frames of video) is unexpectedly occluded. Occlusion can involve freezing the video at the occlusion point (e.g., Johnson & Raab, 2003)—such that the final frame remains visible to the observer—or by replacing the final image with a blank screen (e.g., Ward et al., 2013). In a soccer example, a video of an attacking play may be occluded immediately prior to the player with the ball kicking the ball. The observer/participant is then required to indicate where the ball was about to be kicked (e.g., toward the goal, toward a specific teammate). In other studies, researchers have used liquid-crystal occlusion spectacles to occlude vision during actual task performance (for a review, see Farrow, Abernethy & Jackson, 2005). When their vision is occluded, participants are typically required to continue performing the task under investigation. In both video-based

(1978) asked expert and novice tennis players to watch video clips of an opposing tennis player serving the ball. They occluded the videos at three different occlusion points (i.e., pre-, near-, and post-racket-ball contact). Participants anticipated the ball's landing position on the court, and marked this point on a diagram of a tennis court. Significant differences in anticipation accuracy were observed between the skill groups with expert tennis players being more accurate than their novice counterparts, especially in the pre- and near-racket-ball contact conditions. Using similar temporal occlusion tasks, skill-based differences in anticipation speed and/or accuracy have been demonstrated, for instance, in soccer (e.g., Williams & Davids, 1995), baseball (e.g., Burroughs, 1984), badminton (Abernethy & Russell, 1987), and squash (e.g., Abernethy, 1990), to name but a few. In general, the data suggest that skilled players can anticipate their opponent's intentions earlier and more accurately than their lesser-skilled counterparts. They can often predict the outcome of the situation prior to or immediately after the player with the ball making contact with their foot/racquet/bat—both in small-sided (e.g., 1 vs. 1 badminton) and large-sided games (e.g., 11 vs. 11 soccer).

Subsequent research has used other experimental manipulations, such as spatial occlusion and point-light displays, and recorded process-level data (e.g., eye movements) to provide possible explanations for superior anticipation. This research suggests that experts are better at anticipating because they make effective use of more meaningful postural cue information (e.g., position of the non-kicking foot in a penalty kick as it is placed next to the ball; relative hip-shoulder rotation in tennis prior to racket-ball contact during a serve) and/or interpret differently tactical information conveyed by the relative positions and movements of other players on the field (e.g., Ward, Williams, & Bennett, 2002; Williams & Davids, 1998). Research also indicates that training focused on improving novices' anticipation skills—often using temporal occlusion methods to highlight the cues utilized by expert performers—translates to performance improvement in the field (Fadde, 2006; Gabbett, Rubinoff, Thorburn, & Farrow, 2007; Williams, Ward, Knowles, & Smeeton, 2002).

Although much of the early research required participants to indicate their anticipation using pen and paper, a number of researchers have required participants to respond physically, either by making a forced-choice decision about how to respond to their opponent(s) or by responding physically (e.g., by moving a joystick, or more naturally as they would in a game). The speed and accuracy of participants' decisions and responses have been used to make inferences about their ability to anticipate their opponent's actions, and so provide an indirect measure of

and in situ temporal-occlusion paradigms, participants are required to use the pre-occlusion information to predict what will happen immediately following the point of occlusion.

anticipation skill in the real-world setting. For example, Savelsbergh, Williams, Van Der Kamp, and Ward (2002) required soccer goalkeepers to respond to a video-simulated penalty kick by moving a joystick in the direction they would dive in order to save the penalty. Expert goalkeepers responded more accurately, albeit slightly later than novice goalkeepers. Subsequent analysis of the joystick trajectories indicated that they also made fewer corrective movements of the joystick. In other words, once the experts had selected the direction of the penalty kick, they were less likely to change their mind (and correct their action) as more information became available. Similar results have been found, for instance, in boxing where participants were required to use a joystick to decide which location of the body to protect with a block (e.g., head, chest) in response to an attacking opponent (Ripoll, Kerlirzin, Stein, & Reine, 1995).

Intervention phase decision making. Rather than assessing and/or anticipating what an opponent will do next, a handful of researchers have examined the analogous question in the intervention phase of decision making: When *I* am in possession or control, what will/should *I* do next (e.g., to which location should I serve on the tennis court; to whom should I kick the soccer ball)? In these situations, participants are typically asked to select from a predetermined (and typically limited) set of actions they could perform. Unlike the anticipation studies, where there is a known outcome (e.g., the direction of the soccer penalty kick, the landing point of the tennis serve), the quality of the outcome is assessed subjectively by a panel of experts. For example, Helsen and Pauwels (1993) employed an innovative, representative task to examine decision-making skill in soccer. Expert and novice soccer players responded to a variety of video-based soccer scenarios (i.e., free kicks, penalty kicks, dribbling, and passing situations), presented from the perspective of an on-field player who was involved in the developing situation but who was not yet in possession of the ball; participants adopted this perspective as their own. Each simulation ended with an on-screen teammate passing the ball toward the participant, at which point the video froze. When the on-screen ball was played to the participant, they were required to execute a tactical decision by kicking an actual ball—placed at their feet prior to the simulation—toward a specific teammate/location on the frozen image. The quality and speed of participants' decisions were recorded. Expert soccer players made better decision more often and more quickly than their novice counterparts.

More recently, Gorman, Abernethy, and Farrow (2013) investigated the decision making of skilled and less-skilled basketball players. Players viewed video clips of typical basketball gameplay (i.e., dynamic scenarios) and still-image pictures from a video clip (i.e., static scenarios). The participants were required to adopt the

perspective of the player with the ball and select a response from a limited set of alternatives: pass, dribble to the right, dribble to the left, or shoot. Players' responses were compared to those of an expert panel. For both static and dynamic tasks, skilled players made significantly better decisions than less-skilled players. This growing body of research suggests that both anticipating the situational outcome and selecting the best course of action are useful tasks for differentiating between skill groups.

Option Generation Strategies Supporting Skilled Assessment and Intervention Phase Decision Making

Although some sport-related research has provided a process-level explanation of how individuals successfully anticipate situational outcomes and/or decide on an effective course of action, the resulting explanations and subsequent 'theoretical alignments' have often been exercises in postdiction. Moreover, such explanations are rarely subjected to subsequent empirical test to provide opportunity for falsification. Instead, more post hoc explanation often follows. Only a handful of researchers have tested a priori predictions, generated from extant theory, about how individuals operate in the assessment and intervention phases of decision making in dynamic and complex sporting environments (e.g., Johnson & Raab, 2003; Raab & Johnson, 2007; Ward et al., 2013). Moreover, when non-sports researchers have generated relevant theory from empirical data that is then subjected to empirical test, they have usually ignored the process by which decision makers in complex environments actually generate alternatives from which to choose. This process, henceforth referred to as *option generation*, is considered critical to naturalistic decision-making in a number of dynamic domains (Klein, 1993). Theories of option generation have provided testable hypotheses and process-level explanations of skilled performance during the intervention (Johnson & Raab, 2003) and assessment (Ward et al., 2013) phases of decision making within sport situations.

The take-the-first heuristic. Johnson and Raab (2003) employed an option-generation task in which participants viewed video clips of typical handball play from the perspective of the team with the ball. The clips froze unexpectedly at a critical decision point, leaving the last video frame on screen. The participant's task was to respond, from the perspective of the player with the ball, by generating first an intuitive course of action that they could perform next (e.g., pass the ball to a specific teammate, shoot the ball at goal), then generating all subsequent options that came to mind, and finally selecting the best course of action from amongst those generated. Eighty-five intermediate-skill handball players completed 31 such trials. An expert panel rated the quality of all of the options generated by participants in each of the scenarios presented. Participants generated relatively few options (e.g., 2–3), and the first option generated was of higher quality than subsequently generated options.

To explain this pattern of results, Johnson and Raab (2003) presented the Take-The-First (TTF) heuristic. According to TTF, a response option is activated in memory based on an association with the environmental structure (e.g., a pattern of gameplay in handball). As activation spreads, other workable, but lower-quality, options are generated. Consequently, increasing the number of options that one generates increases the number of lower quality options. Therefore, the likelihood that one can select a high quality option as the final choice is decreased—because a choice would be made from a larger set of options that, on average, were lower in quality. Johnson and Raab predicted that the total number of options generated would be negatively related to the final decision quality. The empirical evidence supported their hypothesis. To establish the extent to which skill level affects the option-generation process, Raab and Johnson (2007) employed the same paradigm with expert, near-expert, and non-expert handball players (total $n = 69$). They found that compared to near-expert and non-expert players, the experts selected final options of higher quality, suggesting that as one acquires more skill their performance aligns with the predictions of TTF more closely. Additionally, during a similar study in handball, Laborde and Raab (2013) found that experts generated significantly fewer options than near-experts, though not fewer than non-experts.

The findings from tests of TTF are consistent with the Recognition Primed-Decision (RPD) model of rapid decision making (Klein, 1989; 1993), which asserts that skilled performers typically generate a workable solution quickly (and first among alternatives), and better performance is associated with generating fewer options. Of particular relevance to this paper, Klein and Peio (1989) suggested that RPD could also describe performance during the assessment phase of decision making. To test this claim, the researchers presented skilled and less-skilled chess players with mid-game chessboard configurations and asked them to generate potential moves that could be made against them, and to anticipate the opposing player's actual next move. Skilled players (a) generated few options—fewer than less-skilled players (albeit not at a significant level), (b) generated the actual next move to be taken by their opponent as their first option more often than less-skilled players, and (c) were more accurate than less-skilled players at predicting oppositional moves. Based on these data (and extrapolating from TTF) one might predict that superior performance during the assessment-phase of decision-making would be supported by intuition, where the first option generated would be superior to subsequently generated options—particularly among highly skilled players—and that generating more situational options would be detrimental to performance.

Long term working memory theory. In a review of research on expert performance, Ericsson and Kintsch (1995) proposed an alternative perceptual-cognitive mechanism for facilitating expert performance in complex and

dynamic domains to that described by RPD and TTF. As one form of retrieval structure, they suggested that experts construct a situational model that indexes and organizes information in long term memory and that integrates this with incoming environmental information on the fly. The result is a constantly updated situational model that facilitates access to multiple, relevant assessment and intervention alternatives (provided multiple relevant options exist in the environment). Ericsson and Kintsch called this mechanism Long Term Working Memory (LTWM) skill. Using an option generation task analogous to that used by Johnson and Raab (2003; Raab & Johnson, 2007), Ward et al. (2013) tested a series of hypotheses derived from LTWM to further investigate LTWM claims about expert skill at cognitively representing relevant alternatives during situational assessment phase of decision making. First, consistent with TTF and RPD—and any activation by association-based model of skill in dynamic tasks—they predicted that experts would generate a good option first and few options in total. However, rather than predict that the TTF and RPD claims would extend to the assessment phase of decision making (e.g., total number of options would be negatively correlated with performance; cf. Johnson & Raab, 2003), Ward et al. generated predictions regarding the number of task-relevant and task-irrelevant options, as opposed to the total options. Task-relevant options in assessment trials were deemed to be those threatening options (available to the opposing attacking team) that a (defensive) player should be aware of at that moment if on the field. Examples of task relevant options available to an opposing soccer player with the ball in a specific situation might include them passing to a specific teammate, continuing to run with the ball before crossing it in to the penalty box, or shooting directly at goal—all of which would pose some degree of threat to defense. This example is illustrated in Figure 1. Dark arrows indicate task-relevant options; light arrows indicate task-irrelevant options. The task-relevant options are representative of the relevant decision alternatives described by Ericsson and Kintsch (1995).

In the research conducted by Ward et al. (2013), task-relevant options were identified by a panel of expert soccer coaches. All options generated by participants that were not identified by the expert panel as task-relevant were deemed task-irrelevant (i.e., not threatening to the defense in each specific situation). Ward et al. (2013) predicted that the number of task-relevant situational options would be positively correlated with the ability to anticipate the opponent's next move (i.e., anticipation accuracy)². Conversely, they predicted that the number of task-irrelevant options would be negatively correlated with anticipation accuracy. Although proponents of TTF

² This is the assessment phase analogue of selecting the best move for oneself in the intervention phase of decision making.

would likely agree that task-relevant options are generated earlier in the decision process, and the first option is often task-relevant, they did not make specific predictions regarding the numbers of task-relevant and -irrelevant options and anticipation accuracy (during assessment) or final decision quality (during intervention). Instead, the predictions of TTF focus on the total number of options and final decision quality.

In three experiments conducted by Ward et al. (2013), skilled- and less-skilled soccer players watched short (i.e., 5–10 s) video clips of typical attacking soccer play from a defensive perspective. Similar to Johnson and Raab's (2003) method, the clips ended unexpectedly at a critical decision moment—immediately prior to an opposing player with the ball performing an action (e.g., shooting toward goal, passing to a teammate). As a defensive player, participants then generated the situational options heeded at that moment (i.e., those options they were concerned that the opposing attacking player might perform next). Participants responded by marking each option on paper on a two-dimensional scaled depiction of a soccer field (from the same perspective as the last video frame). They were also required to indicate which option they thought would actually occur next (i.e., the *anticipated* option), and ranked the options in terms of the threat level posed to the defense. Consistent with predictions from both perspectives, skilled participants were better at anticipating the outcome, generated higher quality options as their first option more often than less-skilled players, and generated only a few options (i.e., 2–5). However, contrary to TTF, the total number of options generated by the skilled and less-skilled groups did not differ significantly. Interestingly though, skilled participants generated more task-relevant and fewer task-irrelevant options than less-skilled participants. Moreover, the numbers of task-relevant and task-irrelevant options generated were positively and negatively related, respectively, to anticipation accuracy. In two of their experiments, Ward and colleagues had participants respond to some trials with perceptual cues (i.e., the last frame of action available on screen as a freeze frame) and other trials without perceptual cues. They observed stronger effects when participants were performed the task without perceptual support (i.e., from memory). Although the performance of participants was consistent with the sentiment of TTF and RPD (i.e., generating few options, a good one first) there was no evidence that generating more options hindered performance—as predicted by TTF. Rather, the ability to generate an accurate situational representation (i.e., containing the available task-relevant options prioritized at encoding, but not task-irrelevant options), as described by LTWM, provided a better explanation for superior performance in these specific situations than the use of a simpler intuitive heuristic.

It would seem logical to suggest, based on the above studies, that the TTF heuristic might provide the best explanation for skilled performance during intervention-phase decision making. Meanwhile, the development of situational representations of the type described by LTWM theory may provide the best explanation for skilled performance during assessment-phase decision making. Accordingly, in this part of the experiment, we examined skilled and less-skilled soccer players' performance on an option generation task. The two theoretical perspectives—LTWM and TTF—permit different predictions to be made with regard to the relationship between performance and the number and type of options generated. For more information about the specific differences between mechanisms the interested reader is referred to Ward et al. (2013). Although other theoretical differences exist, the nature of this relationship in both the assessment- and intervention-phases of decision making is the main focus of our research.

However, subtle differences in methodologies used across studies, including the time available to respond and the presence/absence of perceptual cues when performing experimental tasks may preclude such a conclusion without additional research. Recall that each of the claims by Johnson and Raab (2003) and Ward et al. (2013) were developed to account for performance in time constrained and dynamic domains. Yet, to date, and rather surprisingly, no studies have experimentally manipulated time constraint on option-generation performance in a complex and dynamic domain. Johnson and Raab (2003) allowed participants access to a frozen image on screen (i.e., the final frame of the handball videos) for 45 seconds. Raab and Johnson (2007) reduced time permitted to 6 seconds. Ward et al. (2013) did not allow any time to view the final image on-screen during their occlusion condition, but response time was unlimited. Given the emphasis of these models on generating a satisficing option quickly, in-situ, we explored the effects of time constraint on option generation behavior when performing this task without perceptual cues (i.e., without a frozen image of the last video frame from the test stimuli). In light of the methodological differences across previous studies, our inclusion of a time constraint manipulation was meant to further explore the bounds of the extant claims. Accordingly, hypotheses based on time-constraint are somewhat exploratory in nature but grounded in extant theory.

Predictions Regarding Performance

Compared to less skilled players, we expected skilled players to more accurately anticipate the situational outcome during assessment and select the best criterion option during intervention. Based on the research on performance under pressure (e.g., Suss & Ward, 2010; Gray 2004), we also speculated that skilled players would handle time constraint better than less-skilled players. Accordingly, we expected a significant Time (constrained, not

constrained) \times Skill (skilled, less-skilled) interaction effect on performance (i.e., anticipation accuracy during assessment; selection of the best criterion option during intervention). In other words, we expected the performance of less-skilled players to break down under time constraint, but not the performance of skilled players.

Predictions Regarding Number of Options Generated

In line with TTF and LTWM, we expected that few options (e.g. 2–3) would be generated during both assessment and intervention. However, during assessment—and consistent with Ward et al. (2013)—we expected to observe a Skill (skilled, less-skilled) \times Information Type (task-relevant, task-irrelevant options) interaction. Specifically, we expected that skilled participants would generate more task-relevant, and fewer task-irrelevant, options than less-skilled participants. Given the limited amount of time available to respond under pressure, we speculated that these effects would be affected by time constraint and we would observe an Information Type \times Skill \times Time interaction in the assessment phase. We also expected an Information Type \times Time interaction, such that participants reduce the number of task-irrelevant, but not the number of task-relevant options generated when time constraint is implemented. This expectation is in line with all models of activation by association (e.g., TTF, LTWM). To explore the claims of TTF during assessment, we speculated that a main effect of skill on the number of options generated may be observed (see Laborde & Raab, 2013, Klein & Peio, 1989). If skilled players generated fewer options in total than less-skilled players during assessment trials, this would lend some support for the use of a TTF-type strategy.

During intervention trials, we speculated that skilled participants would generate fewer options in total than less-skilled participants (see Klein & Peio, 1989). Although this is not a direct prediction of TTF (but is a prediction of RPD), the generation of fewer options by skilled players (who performed better on average) in comparison to the less-skilled players would be consistent with the sentiment of TTF. For instance, Laborde and Raab (2013) found that experts generated fewer options than near-experts in handball. We also expected that the total number of options generated would be reduced when under time constraint (i.e., Skill \times Time interaction). To explore the claims of LTWM during intervention trials, we also included information type as a within-subject variable (as we did in the assessment phase) and tested for a three-way interaction (Information Type \times Skill \times Time). If skilled participants generated fewer task-irrelevant, but *more* task-relevant options than less-skilled participants in either time condition, the resulting two-way interaction (Information Type \times Skill) would offer some support for LTWM-type mechanisms during intervention.

Predictions Regarding the Relationship between Option Generation and Performance

Following Ward et al. (2013), we predicted that the number of task-relevant and task-irrelevant options in the assessment phase would be positively and negatively related, respectively, to anticipation accuracy. In line with Raab and Johnson (2007) and Johnson and Raab (2003), during intervention, we expected to observe a negative relationship between the *total* number of options generated and decision quality. The effect of time pressure on these relationships was also explored. During assessment trials, we explored the possibility that time constraint might result in a shift from a LTWM- toward TTF-based strategy (i.e., emergence of a relationship between the total number of options and anticipation accuracy), given the primary emphasis of TTF on generating an immediate satisficing option. During intervention trials, we also explored the very tentative possibility that time constraint could result in a shift from a TTF- toward a LTWM-based strategy (i.e., the emergence of a relationship between task-relevant/-irrelevant options and selection of the criterion best option).

Method

Participants

Skilled participants were 19 male NCAA Division I soccer players. The skilled group averaged 19.78 ($SD = 1.56$) years of age and 11.56 ($SD = 3.36$) years of experience playing soccer. Less-skilled participants were 17 male recreational-level soccer players. The less-skilled group averaged 20.24 ($SD = 1.86$) years of age and 8.06 ($SD = 6.13$) years of experience playing soccer. The skilled participants were paid USD \$20.00 for completing the experimental tasks; the less-skilled participants received university course credit. The data collected from the 17 players in the less-skilled group was part of a larger data set ($n = 21$; 4 females, 17 males) that were presented in another manuscript (see Belling, Suss, & Ward, 2013). We excluded the four less-skilled females to ensure that the groups contrasted here were comparable in regard to gender.

Materials and Procedures

Video simulations ($n = 30$) were created using live soccer match footage. The video footage featured national-level, inter-academy, 18-year old players engaged in competitive soccer match play filmed from an elevated angle above, and behind, the goal. This viewpoint has been shown to be effective for discriminating between skill groups (e.g., Mann, Farrow, Shuttleworth, Hopwood, & MacMahon, 2009; Ward & Williams, 2003). Each video clip was 5–10 seconds in duration. After a brief display of buildup play (e.g., passing, dribbling), the clips were unexpectedly occluded immediately prior to a critical decision moment (see Ward et al., 2013). At this

point, an occlusion image appeared that displayed the field lines (e.g., boundaries, half-pitch mark, and goal box) and the location of the ball on a blank white screen (see Figure 2A); players' locations were omitted. Of the 30 videos created, three each were used assessment-phase and intervention-phase practice trials. The remaining 24 videos were used as test trials for both assessment- and intervention-phases of decision making. Response sheets were used to record option generation. The response sheet for a given trial comprised a copy of the occlusion image displayed on the screen at the point of occlusion, printed onto white, letter-size (216 mm x 279 mm) paper.

Participants viewed the video simulations that were presented using a high-definition video projector. The image was displayed on either a reflective-painted white wall or white video projector screen and was approximately 249 cm wide by 158 cm high. Participants (ranging from 1–6 per session) sat at desks approximately 180 cm from, and facing, the projected image. A video camera above and behind each participant was used to film their response sheet during the trials. This allowed the experimenters to check whether participants were following instructions during the testing session, and to review—post testing—the sequence in which participants generated options. When administering the test to multiple participants, participants were seated such that they could not observe one another's responses; they were not permitted to communicate with each other during the experiment. At the moment of occlusion on-screen, participants were instructed to mentally dump the options that came to mind onto the paper, without filtering their responses. They did this by drawing options onto the response sheets using a simple notation scheme (see Figure 1B for an example). Xs represented defending players, Os represented attacking players, and arrows were used to indicate ball and player movement, as well as the targeted recipient of a pass. During the assessment-phase trials, participants were instructed to envision themselves as a defending player and generate the course(s) of action that the opposing attacker might perform next—beyond the occlusion point. During the intervention-phase trials, participants were instructed to envision themselves as the player with the ball and generate the course(s) of action that they would be considering at the point of occlusion—as if they were on the field. Half of the intervention ($n = 6$) and assessment ($n = 6$) phase trials were non-time-constrained, while the other half were time-constrained. During non-time-constrained trials, participants had unlimited time in which to generate (i.e., draw) their options on the response sheet; in time-constrained trials, participants were given 10 seconds to generate options. The experimenter used a stopwatch in order to implement the time constraint.

After participants recorded their options on the response form, they completed—without time restriction—two rating tasks. For each option generated during an assessment trial, participants indicated how likely it was that

the opposing player would choose that option and how concerning they felt that option was to their defense. For each option generated during an intervention trial, participants indicated how likely it was that they would choose that option if on the field and how good they felt that option was for them to perform, given the situation. All ratings of likelihood and concern/quality were completed using a Likert scale that ranged from 0 (*not at all likely/concerning/good*) to 10 (*very likely/concerning/good*). Participants were instructed to make clear which *single* option they felt as *most* likely and *most* threatening/best (i.e., tied ratings were not allowed).

Scoring and Coding

During assessment trials, the option with the highest likelihood rating was taken to be the participant's *anticipated* option. During intervention trials, the option with the highest likelihood rating was taken to be the participant's *chosen* option. During assessment trials, if a participant's anticipated action matched what actually occurred next (i.e., after the point of occlusion), the trial was considered to be anticipated accurately. If any other option was rated as the most likely option, the trial was considered to be anticipated incorrectly. Similarly, during intervention trials if a participant's chosen option matched the criterion best option, the trial was considered to be selected correctly. If any option other than the criterion best option was rated as the most likely, the trial was considered incorrect. The criterion best option was determined by an expert panel (see below). Anticipation accuracy and selection of the criterion best option were used as performance measures for assessment-phase and intervention-phase decision making, respectively. Participants' overall *anticipation score* during assessment was calculated as the sum of correctly-anticipated time-constrained ($n = 6$) and non-time-constrained ($n = 6$) trials. Similarly, participants overall *selection score* during intervention was calculated as the sum of correct selections in the time-constrained ($n = 6$) and non-time-constrained ($n = 6$) trials. Anticipation and selections scores, therefore, could range from 0 (worst) to 12 (best).

Two subject-matter experts (SMEs) coded participants' responses and determined the task-relevance of each option. SME1 had 18 years of competitive soccer playing experience and one year as team captain at the NCAA Division III collegiate level. SME1 also had one season of coaching experience with a women's collegiate club-level team. SME2 had 17 years of competitive soccer playing experience, including two seasons as a collegiate club-level captain. SME2 also had coaching experience with a youth soccer club for two seasons. In addition to their experience as players and coaches, the SMEs developed familiarity with the stimulus trials by watching each video multiple times with and without occlusion, at normal speed and in slow motion, and also viewing the subsequent

sequence of play. The SMEs determined the task-relevant and task-irrelevant options for each trial before being given access to the data collected.

Participants' responses were categorized based on the action, direction, and location of players and ball movement. The final frame in each trial was divided into zones (e.g., zone A, zone B, etc.). For example, a pass from Player 1 to Player 5 in zone D would be categorized along with other responses that shared those functional and spatial features, whereas a pass from Player 1 to Player 2 in zone D would be grouped as different responses. Just the same, a pass from Player 1 to Player 5 in zone A would be grouped as a different response. SME1 coded all participants' responses. SME2 coded a portion of all responses (approximately 10%). During assessment trials, inter-rater agreement was 85% (Cohen's Kappa = 0.82) in terms of categorical coding. During intervention trials, inter-rater agreement was 82% (Cohen's Kappa = 0.66). The SMEs discussed options they disagreed on and reached a consensus as to its correct categorization. In addition to functional coding, the SMEs generated the task-relevant options for each trial. Again, each option was composed of relevant players, actions, directions, and locations. The task-relevant options generated by the SMEs were considered to be those that a defender should be concerned about (for assessment) and those that an attacker should be considering (for intervention). Agreement regarding task-relevant options was 97% (Cohen's Kappa = 0.96) across all trials. The SMEs also determined the criterion best option for all intervention trials.

Analysis and Results

Performance

Two-way ANOVAs were used to test for the effect of time constraint, skill, and interaction effects on anticipation accuracy (assessment) and selection of best course of action (intervention).

Assessment. During the assessment phase, we observed a significant main effect of skill on anticipation accuracy, $F(1, 34) = 10.21, p = .003, \eta^2_p = .23$. The skilled group ($M = 4.21, SD = 1.62$) anticipated more outcomes correctly than the less-skilled group ($M = 2.65, SD = 1.27$). However, a main effect of time constraint was not observed, $F(1, 34) = 0.08, p = .79, \eta^2_p = 0.002$. Recall we hypothesized that a skill effect on performance would interact with time constraint. However, a Skill \times Time Constraint interaction effect was also not observed, $F(1, 34) = 0.006, p = .940, \eta^2_p < .001$. Skilled participants performed better than less-skilled participants regardless of time constraint.

Intervention. During the intervention phase, we observed a significant main effect of skill on performance $F(1, 34) = 10.53, p = .003, \eta^2_p = .24$. The skilled group ($M = 4.16, SD = 1.54$) selected the criterion best option for more trials than the less-skilled group ($M = 2.36, SD = 1.80$). Again, a main effect of time constraint was not observed, $F(1, 34) = 0.55, p = .460, \eta^2_p = .02$, and counter to our hypothesis, a Skill \times Time Constraint interaction effect was not present, $F(1, 34) = 0.30, p = .590, \eta^2_p = .01$.

Number of Options Generated

Assessment. Factorial ANOVA was used to detect effects of skill and time constraint and information type (task-relevant, task-irrelevant options) on number of options generated during the assessment phase. Recall that we expected a three-way interaction of Skill \times Time Constraint \times Information Type, such that 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. The hypothesized three-way interaction effect was not observed, $F(1, 34) = 0.56, p = .460, \eta^2_p = .02$. However, an Information Type \times Skill interaction effect was observed, $F(1, 34) = 21.57, p < .001, \eta^2_p = .39$. Skilled players generated more task-relevant and fewer task-irrelevant options than less-skilled players, regardless of time constraint (see Table 1). Counter to the hypothesis, we did not observe an Information Type \times Time Constraint interaction effect, $F(1, 34) = 0.49, p = .490, \eta^2_p = .01$. In other words, the hypothesized selective reduction in information type in response to the implementation of time constraint was not observed during assessment trials. Instead, we observed only a main effect of time on number of options generated, $F(1, 34) = 10.09, p = .003, \eta^2_p = .22$. Participants reduced both information types in response to time constraint. Lastly we explored our speculation that experts would generate fewer options in total during assessment. Accordingly, we also tested for a main effect of skill on number of options generated; a significant effect was not observed, $F(1, 34) = 0.43, p = .520, \eta^2_p = .01$. Skilled players did not generate fewer options in total than less-skilled players during assessment trials.

Intervention. During intervention, we expected to observe an interaction between skill and time, and a main effect of skill on number of options generated. Although the interaction was not significant, the skill main effect was, $F(1, 34) = 6.17, p = .020, \eta^2_p = .15$. Skilled participants generated fewer options in total than less-skilled participants (see Table 2). 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 \times Time Constraint \times Information Type). Although the 3-way interaction was not

significant, $F(1, 34) = 0.83, p = .370, \eta^2_p = .02$, the hypothesized Information Type \times Skill effect was observed, $F(1, 34) = 24.20, p < .001, \eta^2_p = .42$. Skilled participants generated significantly more task-relevant and fewer task-irrelevant options than less-skilled participants (see Table 2). Moreover, the type of information generated under time constraint affected all participants, $F(1, 34) = 5.11, p = .030, \eta^2_p = .13$. Participants reduced only task-irrelevant options when time constraint was implemented, resulting in a selective—rather than a general—reduction in the number of options generated.

Relationship between Option Generation and Performance

Multiple regression analyses were used to investigate the relationship between option generation and performance during both assessment and intervention phases of decision making. We followed up these analyses and partitioned the variance using correlation (and Bonferroni correction where necessary) since the specific hypotheses were stated in correlational terms.

Assessment. 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. This analyses revealed that together, the numbers of task-relevant and task-irrelevant options explained a significant amount of the variance in anticipation accuracy, $R^2 = .41, F(2, 33) = 19.65, p < .001$. 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). As hypothesized—and consistent with Ward et al. (2013)—the number of task-relevant options was positively related to performance ($r = .61, p < .001$), and the number of task-irrelevant options was negatively related to performance ($r = -.32, p = .030$). These relationships remained—and in the same directions—when we analyzed data from the time-constrained and non-time-constrained conditions separately.

Recall that TTF posits a negative relationship between the *total* number of intervention options generated and performance, and that we extended this hypothesis to the assessment phase of decision making. Contrary to this hypothesis, a correlational analysis revealed a small positive, but not significant, relationship between the number of assessment options and anticipation accuracy ($r = .12, p = .510$). This relationship did not change when we analyzed data from the time-constrained and non-time-constrained conditions separately.

Intervention. In accordance with TTF, we hypothesized a negative correlation between the total number of options generated and performance (i.e., selection of the criterion best option). Across both time-constraint

conditions, the hypothesized relationship was not significant, but trended in line with TTF and our hypothesis ($r = -.24, p = .160$). The relationship did not change when we analyzed data from the time-constraint conditions separately. Correlations for the data in time-constrained ($r = -.21, p = .210$) and non-time-constrained ($r = -.18, p = .300$) conditions were not significant.

In order to explore 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, $R^2 = .36, F(2, 33) = 10.77, p < .001$. As in assessment, the variance was partitioned using correlations with Bonferroni corrections ($\alpha = 0.025$ level). The numbers of task-relevant ($r = .43, p = .009$) and task-irrelevant ($r = -.56, p < .001$) options were significantly related to performance in the directions associated with the LTWM-based claims of Ward et al. (2013).

Discussion

Consistent with the body of research on anticipation and decision making in dynamic situations, skilled participants outperformed less-skilled players during both assessment (by anticipating outcomes) and intervention (by selecting the criterion response option) trials of the representative sport task used in this research (e.g., Abernethy, 1990; Abernethy & Russell, 1987; Gorman et al., 2013; Helsen & Starkes, 1999; Johnson & Raab, 2003; Raab & Johnson, 2007; Ward et al., 2013; for a review, see Mann, Williams, Ward, & Janelle, 2007). However, time constraint did not interact with performance during assessment or intervention. We speculate that this may be because time constraint is relatively common in the domain of soccer and therefore does not degrade performance or that the manipulation was not stringent enough to have a severe effect on performance (but see option generation results, in particular during intervention). It also is possible that written responses, as opposed to oral, affected the relationship between performance and time constraint. Future research should investigate the effect of time constraint on performance across varying response formats.

The number and type of options generated were comparable to the data observed by Ward et al. (2013) during both assessment and intervention. Skilled participants generated more task-relevant and fewer task-irrelevant options than less-skilled players suggesting the use of a LTWM-type strategy during assessment and intervention. According to LTWM, as skilled decision-makers develop, their encoding of situational information becomes more task-relevant and elaborate, resulting in a larger number of the available task-relevant options and fewer task-irrelevant options forming part of the representation. Although a three-way interaction with time constraint was not

observed in assessment, the effects of time constraint were different in the two decision phases. During assessment, there was no effect of time constraint on the numbers of task-relevant and task-irrelevant options generated. Counter to our expectations we did not observe a general reduction in the number of options generated with time constraint during assessment that might have reflected a shift from a LTWM towards use of a TTF strategy. Moreover, although the correlation between the total number of options generated and anticipation accuracy was in the direction hypothesized by Johnson and Raab (2003), albeit not significant, we observed significant positive/negative relationships between task-relevant/irrelevant options generated, respectively, and anticipation accuracy. These data are consistent with the hypotheses derived from LTWM by Ward et al. (2013) and as we hypothesized here. Moreover, these relationships were not considerably different across time constraint conditions.

During intervention, while the hypothesized three-way interaction between skill level, information type, and time constraint was not observed, skilled participants still generated more task-relevant and less task-irrelevant options than less-skilled players. Although there was not an interaction effect between time constraint and skill on performance, the manipulation of time constraint did result in a selective reduction in the number of task-irrelevant options generated by all participants. These data are consistent with the prescription from TTF that participants should *Take-The-First* during the intervention phase of decision making when under time constraint. However, the total number of options generated was not significantly correlated with decision quality (albeit in the hypothesized direction) as would be expected by TTF. Since both the number of task-relevant and -irrelevant options were positively and negatively related, respectively, with decision quality, collectively, these data suggest the use of a strategy that is more consistent with LTWM during intervention rather than a TTF-type strategy as we had predicted.

In conclusion, our results provide stronger, albeit tentative, support for the use of LTWM-based mechanisms during the assessment and intervention phases of decision making in the types of representative sports tasks used in this experiment. The generation of more task-relevant, and fewer task-irrelevant, options was associated with higher skill (as seen in the factorial ANOVA) and higher performance (as seen in the regression). This is consistent with recent explorations of TTF, such as Laborde and Raab's (2013) description of their results that, "...suggest that variables related to quality are more robust for distinguishing expertise level" (p. 349). Similarly, we found that organization of the options generated by quality (i.e., task-relevant vs. task-irrelevant) is more distinguishing of expertise level. The results are also consistent with the prescription of TTF during

intervention (i.e., take the first option) and the general sentiments of intuitive heuristics like TTF and RPD (e.g., better options generated first, generate few options, and reduce the number of irrelevant options generated when it matters, for instance, under time constraint). More research is needed to further delineate the bounds of the usefulness of these heuristics and strategies, especially under time constraint. In particular, research that examines the serial position and quality of each option during both assessment- and intervention-phases of decision making is needed in future work. Another potential limitation of the current research is the use of only moderately skilled SMEs when determining the task-relevant and task-irrelevant options for each trial. However, the data suggest greater convergence between the experts and SMEs than between novices and SMEs. One might predict that the skill effects observed in this research may be more profound if SMEs of the same (or perhaps higher) skill level as the high-skill group were to establish the criterion response. Although this is substantiated by Ward et al.'s (2013) data, further research is needed to address this issue.

The results have implications for training; however, we add the caveat that any single strategy is unlikely to be effective in all situations and that one of the most important features of training for decision making is that it should be context sensitive (see Hoffman, Ward, Feltovich, Dibello, Fiore, & Andrews, 2014). During intervention-phase decision making in the types of situations investigated here, instead of focusing on a *general* reduction in option generation, training should focus on awareness of the of task-relevant options available in the environment/situation, and on selectively reducing the number of task-irrelevant options generated. A quickly-generated, intuitive response appears to be superior to subsequently generated options, especially when under time pressure, but we are less certain that this intuitive representation and response should comprise a single option when multiple relevant options are available in the environment/situation.

During assessment-phase decision-making in the types of situations investigated here, training should focus on building a better representation of the ecological structure by promoting—during training—an awareness of the task-relevant options available in the environment/situation. Over time, this should facilitate development of organized retrieval structures that permit direct access to task-relevant options, and the ability to generate as many of the task-relevant options as are available in the environment/situation, while reducing only task-irrelevant information. Given that the total number of assessment options was not related to performance or skill, we do not see a benefit in attempting to reduce the total number of options to a single, intuitive response. Instead, we recommend that training focuses on the situational-model-building process described by LTWM (see Ericsson &

Kintsch, 1995), such that naturalistic decision makers focus on maintained access to the situational outcome *and* relevant assessment alternatives (see Ward et al., 2013; cf. Klein & Peio, 1989).

EXPERIMENT 1B

Despite the utility of the option generation task (used in Experiment 1A) to differentiate between skill groups and provide insight into the strategies used in each phase of decision making—both with and without time constraint (at least within the representative soccer scenarios investigated)—it is unlikely to be adopted as a diagnostic tool in the field because it is very time-consuming to code, and effortful to administer and score. Recall that the second aim of this paper was to create a diagnostic tool for assessing perceptual cognitive skill in soccer that is readily available and usable by domain professionals. We also sought to evaluate this tool's ability to predict cognitive strategy (e.g., LTWM-type option generation), in addition to domain skill. Accordingly, in this experiment we created and evaluated a new, more usable, and online test of perceptual-cognitive skill in soccer: The Online Assessment of Strategic Skill In Soccer (OASSIS). This test examines players' ability to anticipate the actions of soccer players performing in dynamic video scenes, similar to those used in Experiment 1A. The OASSIS is based on the traditional temporal-occlusion paradigm (for a full review, see Williams & Ward, 2007). Information is presented up until a critical decision point before being occluded. At the point of occlusion, participants must select what they think will happen next from multiple predetermined options (as opposed to generating the options as in the option-generation paradigm; see Ward et al., 2013). Following Ward et al. (2013), and to capture more of the ecological structure to facilitate subsequent generalization, three different types of questions were asked in a variety of situations (e.g., passing, dribbling, shooting): What action will be performed next? In which direction will the pass be played? Which player will receive the ball? This type of test is not new—although simple online tests of perceptual-cognitive skill in sport have not yet been developed. However, the examination of construct validity with respect to option generation behavior is new (see below).

To assess the known-groups validity of the OASSIS, we tested its ability to differentiate between two different skill groups in terms of overall anticipation accuracy. We hypothesized that there would be a significant effect of skill on OASSIS performance. Specifically, we predicted that the skilled group would anticipate correctly the outcomes of scenarios significantly more often than the less-skilled group. To assess the predictive validity of the OASSIS we tested the relationship between performance on this test and anticipation accuracy on the option-generation task using in Experiment 1A. Since the OASSIS also measured anticipation accuracy we expected these

measures to be positively, and significantly, correlated. Lastly, to assess construct validity, we examined the relationship between option-generation behavior during assessment on the option-generation task and performance on the OASSIS. We expected these to follow the relationships predicted by Ward et al. (2013) even though the tests and stimuli were otherwise unrelated and independent. More specifically, we expected to observe convergence (i.e., a positive relationship) between the number of task-relevant options generated on the assessment phase option generation task from Experiment 1A and performance on the OASSIS. Similarly, we expected to observe divergence (i.e., a negative relationship) between the number of task-irrelevant options generated and OASSIS performance.

Method

Participants

Rather than independently replicate the findings of option generation task used in Experiment 1A (which has been replicated elsewhere, see Ward et al., 2013), our primary goal was to compare performance across the two tasks (OASSIS; option generation task) and establish validity. Hence, participants from Experiment 1A were recruited to participate in Experiment 1B. Recruitment of participants was conducted via email and was on a voluntary basis. The high- and low- skill groups were formed using the same participants from Experiment 1A. To prevent alpha inflation due to the use of the same participants, our analyses were Bonferroni corrected.

Materials and Procedure

OASSIS. Video simulations were created from the same bank of soccer footage as in Experiment 1A, resulting in new simulations that used similar—but not the same—critical decision moments and occlusion points. As in the assessment version of the option-generation task (Experiment 1A), video simulations ended unexpectedly with an occlusion image denoting the field lines and position of the ball at the end of the clip. In addition to these features, the occlusion image contained performance lines (e.g., the direction a player may run) and ball position to provide explicit multiple-choice options in three types of simulation: action, direction, and pass recipient. For action simulations, participants were presented with three assessment options from which to choose: pass, shoot or dribble the ball. For both direction and pass-recipient simulations, participants were presented with 3 to 4 assessment options marked on the occlusion image (see Figures 3 and 4, respectively). Video simulations were presented to participants via Qualtrics.com, an online survey website using Adobe Flash (www.adobe.com/products/flashplayer.html) to display video files. Immediately after watching each video

simulation, the multiple choice options were presented on-screen as radio buttons. Participants viewed 15 trials of each simulation type.

Participants received an email containing a link to the survey website inviting them to participate. After providing informed consent and checking their computer could display Flash Video, participants completed the OASSIS using a computer with internet access (e.g., at home) within two weeks of completing Experiment 1A. Participants first completed three training trials, one for each simulation type (i.e., action, direction, and pass recipient). Feedback was not provided during training but participants were permitted to watch the clips more than once until they understood the task fully. During test trials, participants watched each video clip once from a defensive perspective, and were instructed to anticipate the action, direction, or pass recipient—depending on the trial type—by selecting one of the presented options. They were instructed to complete each trial as quickly and accurately as possible. After reading and agreeing to these instructions, participants were automatically guided through the 45 test trials, which were presented in random order. A final score was computed by calculating the number of trials answered correctly (maximum score = 45). The score was not made available to participants but they were invited to contact the experimenter for feedback.

Analysis and Results

Option-Generation Measures. To assess construct validity between the OASSIS and the more time-demanding option-generation task, we used data from Experiment 1A. Specifically, the anticipation scores from Experiment 1A were used as measures of assessment performance. Additionally, the numbers of task-relevant and task-irrelevant options generated during assessment-phase decision making of the option-generation task from Experiment 1A were used as measures of option-generation behavior.

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 analysis because their average time fell more than two 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 more than double the average time. In one case, the time taken was below average, such that it would have been impossible for the participant to have viewed each video clip in its entirety. The times for the remaining participants were consistent with having watched each video once and then responding

to each question within a few seconds. Therefore, data from 16 skilled players and 17 less-skilled players were included in the subsequent analyses.

A one-way (between-participants: skilled, less-skilled) ANOVA was used to detect a skill effect on OASSIS performance; Cohen's d was used to estimate effect size. The skilled group ($M = 28.06$, $SD = 4.21$) anticipated the outcome significantly more frequently than the less-skilled group ($M = 24.12$, $SD = 4.94$), $F(1, 31) = 6.06$, $p = .020$, $d = 0.86$. Pearson's correlations were used to investigate whether performance on the OASSIS predicted performance on the assessment version of the option generation task. Since data from the option generation task had been analyzed in the previous experiment we used a Bonferroni correction to adjust the alpha level ($.05/2 = .025$). Anticipation accuracy scores on both tests were correlated ($r = .35$, $p = .040$) but only approached significance when we applied the Bonferroni-corrected alpha level.

Pearson's correlations were also used to investigate whether construct validity could be established by examining the relationship between performance on the OASSIS and the numbers of task-relevant/task-irrelevant options generated during Experiment 1A for assessment trials using the option-generation paradigm. As hypothesized, the number of task-relevant options generated converged with (i.e., was positively related to) performance on OASSIS ($r = .49$, $p < .010$). Additionally, we also hypothesized a divergent (i.e., negative) relationship between the number of task-irrelevant options generated during assessment and performance on the OASSIS. Although a negative relationship was observed, it was not statistically significant ($r = -.22$, $p = .220$). The significance of these correlations was not affected when the Bonferroni-corrected alpha level was applied.

Discussion

Experiment 1B demonstrated partial support for the validity of the OASSIS as a tool for assessing perceptual-cognitive skill in soccer. Although traditional temporal-occlusion tasks have been used in previous research to differentiate between skill groups (see Abernethy, 1990; Abernethy & Russell, 1987), to our knowledge, this is the first time that an online version of this paradigm has been used to capture skill-based differences in anticipation performance. Furthermore, we demonstrated that anticipation accuracy on OASSIS correlates with anticipation accuracy on the more in-depth option-generation paradigm (see Ward et al., 2013). However, this correlation only approached significance when we corrected for multiple contrasts. Subsequent research should further examine these issues using a new and independent sample.

Most importantly, perhaps, this is the first time that construct validity has been demonstrated based on the theoretical predictions about underlying cognitive processes facilitating skilled performance. We expected anticipation accuracy on the OASSIS to converge and diverge, respectively, with the numbers of task-relevant and task-irrelevant options generated on the assessment phase of the option generation task. This was partially supported. Performance on the OASSIS was significantly related to the number of task-relevant options generated (it also trended negatively with the number of task-irrelevant options generated, as we had hypothesized). This in turn provides further support for the use of LTWM-type mechanisms to facilitate skilled performance, as is suggested in Experiment 1A. A limitation of the research was maintained access to many highly skilled players. Future research should examine these relationships among independent and larger samples of skilled and low-skilled players. Moreover, future research should investigate these relationships among even more highly skilled athletes (e.g., professional-level players).

Another limitation of the OASSIS in this research was the use of only assessment-phase stimuli. In the assessment phase, the actual outcome can be used as the correct answer on trials; this simplifies the scoring process, and makes for an easy-to-use, applied tool. An intervention-phase version of the OASSIS, however, would require more complicated scoring; future research should investigate this by creating correct responses to intervention-phase situations as agreed upon by domain experts. Just as has been done in this experiment, an intervention-phase version could be evaluated by establishing the discriminative power between skill groups. Furthermore, the relationship between performance on this test and performance/option-generation on the intervention-phase of the option generation task could be used to assess the validity of an intervention-phase test. Future research should investigate expanding the OASSIS to cover both assessment and intervention phases.

In conclusion, given these limitations, the option generation paradigm (described in Experiment 1A) may be the best predictor of perceptual-cognitive skill in soccer. However, given that it is not likely to be used by domain professionals, Experiment 1B has demonstrated that the OASSIS may be a useful diagnostic tool in its place. The data demonstrate that the OASSIS is a useful tool for assessing perceptual-cognitive skill (i.e., anticipation) in soccer. To our knowledge, this is the first time a tool with such high usability and accessibility has been empirically validated. Although future research is needed to further explore the utility of the OASSIS under different situations (e.g., intervention-phase soccer situations) and for different skill groups (e.g., younger players, professional-level

players), this research suggests the use of the OASSIS for this purpose is promising and further investigation is warranted.

EXPERIMENT 1C

Given that Experiment 1B demonstrated the utility of the OASSIS at discriminating between skill groups, and its predictive and convergent validity relative to the option generation test, our final aim was to compare its predictive power relative to other domain-general cognitive tests. In prior research that has examined the relative contribution of domains-specific skills and domain-general abilities to soccer skill, researchers have generally found that domain-specific skills have more predictive power. For instance, Helsen and Starkes (1999) reported that adult expert and novice soccer players did not differ significantly on many domain-general measures, such as simple reaction time, peripheral reaction time, visual correction time, static acuity, dynamic acuity, mesopic acuity, and horizontal and vertical peripheral range. However, expert soccer players were quicker and more accurate than novices when deciding what to do next in a soccer-related decision task. Likewise, Ward and Williams (2003) found similar results when comparing performance of soccer players ranging from 9- to 18-years old on domain-general and domain-specific tasks. That is, the scores on the domain-specific skills tests—specifically, anticipation and situation assessment—combined to predict skill-group membership. Domain-general ability tests, including static and dynamic visual acuity, stereoscopic depth sensitivity, and peripheral awareness did not contribute to the significant model.

The domain-general tests used in these two, and many other studies were primarily visual-perceptual in nature. In a recent meta-analysis of 20 studies examining the basis for superior visual, perceptual, and cognitive skill across multiple sports, Voss, Kramer, Basak, Prakash, and Roberts (2010) found that domain-general cognitive abilities contributed to a small, but significant skill effect. Their results showed a skill effect for cognitive processing speed ($ES = 0.67, p < 0.05$) and varied attentional paradigms ($ES = 0.53, p < 0.05$), such as the Paced Auditory Serial Addition Task (Gronwall, 1977) and the Eriksen arrow flankers task (Eriksen, 1995). When averaged across a number of studies, skilled athletes scored higher on tests of cognitive processing speed and attention than less-skilled counterparts.

In Experiment 1C, to examine whether the validated OASSIS test was amongst one of the more useful predictors of skill in sport, we sought to address the argument put forth by Voss et al. (2010)—that general cognitive abilities (as opposed to visual-perceptual abilities) are also predictive of skill. More plainly, although previously

observed effects have been noted as being small, skilled athletes tend to demonstrate higher general cognitive ability. We examined whether performance on OASSIS could predict skill group membership, over and above measures of domain-general cognition, specifically those that tap into the constructs identified by Voss and colleagues.

Because Voss et al. (2010) found the largest effects in tests of cognitive processing speed and attention, we selected a robust, and very short, test of domain-general cognition that is significantly related to tests that tap into these constructs—the Berlin Numeracy Test (BNT; Cokely, Galesic, Shulz, Ghazal, & Garcia-Retamero, 2012, see Appendix A). The BNT was designed to measure statistical numeracy and risk literacy and is a brief, robust psychometric measure of domain-general cognition (Cokely, et al., 2012). 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 in soccer, and performance on the OASSIS, require participants to mentally map out actions and players over space and time. For this reason, it is conceivable that a more general measure of spatial skill might also predict soccer skill level. Accordingly, we included the Mental Rotations Test (MRT-A; Vandenburg & Kuse, 1978) (see Appendix B) as a domain-general cognitive predictor in our analyses. This test is a reliable and valid measure of general spatial skill (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Consistent with Voss et al. (2010), we expected that performance on the BNT and MRT-A would explain some of the variance in skill level. However, we also predicted that anticipation accuracy—as measured by OASSIS—would also predict skill group but would explain more variance between skill groups than the domain-general measures.

Method

Participants

The less-skilled group comprised 35 male recreational-level soccer players, 13 of whom also participated in Experiment 1B and 1A. The skilled group comprised the 16 NCAA Division I soccer players from Experiment 1B. Again, in order to prevent alpha inflation due to the use of the participants in previous experiments, our analyses were Bonferroni corrected. Participants were recruited via email and on a voluntary basis.

Materials and Scoring

The BNT (see Cokely et al., 2012) contains four psychometrically validated questions to assess statistical numeracy and risk literacy (see Appendix A), which were formatted in Qualtrics.com survey builder. 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 the procedure described by 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. 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.

The OASSIS, and scoring method, were the same as described in Experiment 1B.

Procedure

Participants received an email containing a link that directed them to the Qualtrics.com survey containing the tests. Participants completed the BNT and then the MRT-A. Participants who had not participated in Experiment 1B then completed the OASSIS. 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 tests, the “back button” was disabled, preventing participants from changing their responses. When participants finished all the tests, they were thanked for their participation and the survey ended automatically. Participants were not able to ascertain feedback or re-do any test questions, though participants were invited to contact an experimenter for feedback on their scores, if they so desired.

Analysis and Results

Since data on OASSIS performance had been analyzed in the previous experiment we used a more conservative alpha level ($.05/2 = .025$). A logistic regression analysis was conducted 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 (β) associated with each variable were compared. Performance on OASSIS was a significant predictor of skill group membership ($\beta = 10.81, S.E. = 4.51, p = .020$). The BNT ($\beta = -0.04, S.E. = 0.04, p = .340$) and MRT-A ($\beta = -0.57, S.E. = 0.40, p = .160$) were not significant predictors of skill group membership.

Discussion

As hypothesized, performance on the OASSIS explained more variance between skill groups than domain-general measures—even the types of variables highlighted by Voss et al. (2010). We take this as evidence that the OASSIS is a more valid predictor of soccer skill group membership than domain-general measures. These results demonstrate that the OASSIS is perhaps the most useful online tool for predicting the skill level of soccer players. Consistent with previous research (see Helsen & Starkes, 1999; Ward & Williams, 2003), our results suggest that domain-specific skill—namely anticipation skill—is more predictive of skill-group membership than domain-general measures of cognition. This is at least the case for the general cognitive measures selected in this research. Although access to highly-skilled players was difficult to secure—as is often the case in expertise research—future research should endeavor to replicate the discriminative power of the OASSIS among new and independent skill groups to provide more substantiation to the claims presented in this manuscript. Again, this was a limitation in this research. Furthermore, future research should also evaluate whether training perceptual-cognitive skill via the OASSIS improves performance on both representative tasks and on the soccer field. Such research would demonstrate the utility of the OASSIS stimuli not only as a predictor of soccer skill, but as a simple and useful platform for training perceptual-cognitive skill in soccer. Lastly, future research should evaluate the validity of OASSIS-type tools in other complex and dynamic domains where the perceptual-cognitive skill is a critical component of skilled performance.

Conclusions

Across a three-part experiment, we offer theoretical contributions to the area of decision-making in complex and dynamic domains—particularly those that utilize option generation to explain superior performance. We also provide an empirically validated and applied tool for capturing perceptual-cognitive skill. Experiment 1A

supported the notion that individuals typically generate few options (i.e., 1–3) when anticipating an opposing player's action, and when—as the player with ball—selecting a course of action to pursue themselves. Additionally, there was a non-significant trend in the data suggesting that generating more options—regardless of their task relevance—was negatively related to performance during the intervention phase (see TTF; Johnson & Raab, 2003). In particular during the intervention-phase, the selective reduction of task-irrelevant information in response to time constraint by participants of both skill groups suggests that performance was facilitated by a mechanism that reduces towards fewer intuitive options. The initial response contained task-relevant options, and subsequently generated options (heeded during additional permitted time) that were mainly task-irrelevant. For this reason, we expect that the TTF heuristic may provide a useful *prescription* to naturalistic decision makers in sport and other complex and dynamic domains (i.e., reducing the generation of additional, and specifically, irrelevant information is associated with higher performance and higher skill). The results showed that these trends were stronger during intervention-phase trials than for assessment-phase trials. However, even during assessment, the generation of fewer task-irrelevant options was associated with higher performance and skill.

Despite support for the prescriptions that are proposed by proponents of TTF, the results from Experiment 1A are more consistent with the hypotheses of Ward et al. (2013) and may provide a better *explanation* of expert performance in this context. The numbers of task-relevant and task-irrelevant options generated were more strongly related to performance and skill—in the directions proposed by LTWM theory (i.e., a positive/negative relationship between the numbers of task-relevant/task-irrelevant options and performance; see Ward et al., 2013)—than simply the total number of options generated (cf. Johnson & Raab, 2003). This may have implications for training. Instead of emphasizing a reduction in the amount of information generated toward one single intuitive option (as may be beneficial according to TTF, or at least extreme views of RPD), emphases should be placed on identifying as many of the relevant and/or threatening decision alternatives in the environment/situation as are available, while reducing the task-irrelevant information. These results are consistent with having a good situational representation during the assessment phase, which will likely facilitate selection of a high-quality response during the intervention phase (see Ward et al., 2011).

In the intervention phase, the skilled group generated significantly fewer options than the less-skilled participants and the total number of options generated trended negatively with performance, which we took as support for a TTF-like mechanism (see Johnson & Raab, 2003; Raab & Johnson, 2007). Only by evaluating

competing hypotheses regarding the direction of the relationship between the number of options generated, their relevance, and performance were we able to draw conclusions regarding which theory—TTF or LTWM (Ward et al., 2013)—better accounted for the data from assessment- and intervention-phase trials. Based on the analyses, we concluded that the data were consistent with the types of mechanisms proposed by LTWM theory. However, our results do not conflict with the *prescription* of TTF during the intervention phase (i.e., that decision-making during the intervention-phase would likely benefit from taking the first option generated).

Experiments 1B and 1C provide support that a new, relatively simple online test of perceptual-cognitive skill in soccer—OASSIS—showed some promise as a diagnostic tool. . Four key findings were observed: Performance on OASSIS (a) differentiated between skill groups and, hence, known-groups validity was demonstrated; (b) was correlated with anticipation accuracy during assessment-phase decision making on the option-generation task—albeit fell marginally below the corrected significance level—hence, provided some tentative evidence for its predictive validity; (c) was correlated with an independent measure of option generation strategy use, providing convergent validity with respect to the underlying cognitive processes supporting performance—and therefore offers some utility in terms of diagnosing strategic deficiencies in perceptual-cognitive skill; and (d) was a better predictor of skill level than domain-general measures of cognition that have been identified previously as characteristic of both general and sports-specific ability. In future research, we plan to replicate these findings with an independent sample, extend the research by developing an intervention-phase decision making version of OASSIS, conduct an item analysis on the OASSIS data and, given its diagnostic capability, develop an adaptive version that maximizes usability and efficiency, thereby increasing its appeal to both the scientific and applied communities.

References

- Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sport Sciences*, 8, 17–34.
- Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science*, 6, 283–319.
- 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., Suss, J., & Ward, P. (2013). Investigating constraints on decision making strategies. In Chaudet, H., Pellegrin, L., & Bonnardel, N. (Eds.) *Proceedings of the 11th International Conference on Naturalistic Decision Making (NDM 2013), Marseille, France, 21-24 May 2013*. (pp. 133–136). Paris, France: Arpege Science Publishing.
- Burroughs, W. A. (1984). Visual simulation training of baseball batters. *International Journal of Sport Psychology*, *15*, 117–126.
- 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.
- Engström, D. A., Kelso, J. A., & Holroyd, T. (1996). Reaction-anticipation transitions in human perception-action patterns. *Human Movement Science*, *15*, 809–832.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Eriksen, C. E. (1995). The flankers task and response competition: A useful tool for investigation of a variety of cognitive problems. *Visual Cognition*, *2*, 101–118.
- Fadde, P. J. (2006). Interactive video training of perceptual decision-making in the sport of baseball. *Technology, Instruction, Cognition and Learning*, *4*, 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*, 330–349.
- Gabbett, T., Rubinoff, M., Thorburn, L., & Farrow, D. (2007). Testing and training anticipation skills in softball fielders. *International Journal of Sports Science and Coaching*, *2*, 15–24.
- 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*, *66*, 2219–2236.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42–54.
- Gronwall, D. M. (1977). Paced auditory serial-addition task: A measure of recovery from concussion. *Perceptual Motor Skills*, *44*, 367–373.
- Haskins M. J. (1965). Development of a response-recognition training film in tennis. *Perceptual and Motor Skills*, *21*, 207-211.

- Helsen, W. F., & Pauwels, J. M. (1993). The relationship between expertise and visual information processing in sport. In J. L. Starkes, & F. Allard, (Eds.), *Cognitive issues in motor expertise* (pp. 109–134). Amsterdam: Elsevier.
- Helsen, W. F., & Starkes, J. L. (1999). A multidimensional approach to skilled perception and performance in sport. *Applied Cognitive Psychology, 13*, 1–27.
- Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D., (2014). *Accelerated expertise: Training for high proficiency in a complex world*. New York, NY: Psychology Press.
- Johnson, J. G., & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes, 91*, 215–229.
- 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*, 231–235.
- Klein, G. A. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research* (Vol. 5, pp. 47–92). Greenwich, CT: JAI Press, Inc.
- 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., & Peio, K. J. (1989). Use of a prediction paradigm to evaluate proficient decision making. *The American Journal of Psychology, 102*, 321–331.
- Laborde, S., & Raab, M. (2013). The tale of hearts and reason: The influence of mood on decision making. *Journal of Sport & Exercise Psychology, 35*, 339–357.
- Mann, D. L., Farrow, D., Shuttleworth, R., Hopwood, M., & MacMahon, C. (2009). The influence of viewing perspective on decision-making and visual search behavior in an invasive sport. *International Journal of Sport Psychology, 40*, 546–564.
- Mann, D. T. Y., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta analysis. *Journal of Sport & Exercise Psychology, 29*, 457–474.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York, NY: W. H. Freeman.
- 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.
- Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied*, 13, 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, 325–349.
- 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.
- Suss, J., & Ward, P. (2010). Skill-based differences in the cognitive mechanisms underlying failure under stress. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54, 1062–1066.
- 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.
- 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.
- 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., 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*, 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.
- Ward, P., Williams, A. M., & Bennett, S. J. (2002). Visual search and biological motion perception in tennis. *Research Quarterly for Exercise and Sport*, 73, 107–112.

- Ward, P., Williams, A. M., & Hancock, P. A. (2006). Simulation for performance and training. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 243–262). New York: Cambridge University Press.
- 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*, 259–275.
- Williams, A. M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research quarterly for exercise and sport*, *69*, 111–128.
- Williams, A. M., & Ward, P. (2007). Anticipation and decision making: Exploring new horizons. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 203–223). Hoboken, NJ: John Wiley & Sons.
- 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*, 259–270.
- Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review*, *15*, 763–771.

	Info. type	Skill group	Time constraint	No time constraint	Both time conditions
Number of options generated	Task-relevant	Less-skilled	4.35 (1.97)	4.71 (2.44)	9.06 (3.80)
		Skilled	6.11 (2.23)	6.84 (2.03)	12.95 (3.27)
		Cohen's <i>d</i>	$d = 0.84$	$d = 0.95$	$d = 1.10$
	Task-irrelevant	Less-skilled	5.59 (2.32)	6.88 (3.95)	12.47 (5.41)
		Skilled	3.47 (2.52)	3.79 (2.78)	7.26 (4.20)
		Cohen's <i>d</i>	$d = -0.72$	$d = -0.90$	$d = -1.08$
	Total	Less-skilled	9.94 (3.01)	11.59 (3.68)	21.53 (6.00)
		Skilled	9.58 (2.91)	10.63 (3.50)	20.21 (6.12)
		Cohen's <i>d</i>	$d = -0.12$	$d = -0.27$	$d = -0.22$

Table 1: Means (standard deviations) of number and type of options generated by the high- and less-skilled groups during assessment-phase trials.

	Info. type	Skill group	Time constraint	No time constraint	Both time conditions
Number of options generated	Task-relevant	Less-skilled	4.88 (2.62)	4.18 (1.98)	9.06 (3.93)
		Skilled	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	Less-skilled	5.53 (2.67)	7.29 (2.93)	12.82 (4.63)
		Skilled	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	Less-skilled	10.41 (3.00)	11.82 (2.88)	22.24 (5.14)
		Skilled	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

Table 2: Means (standard deviations) of number and type of options generated by the high- and less-skilled groups during intervention-phase trials.

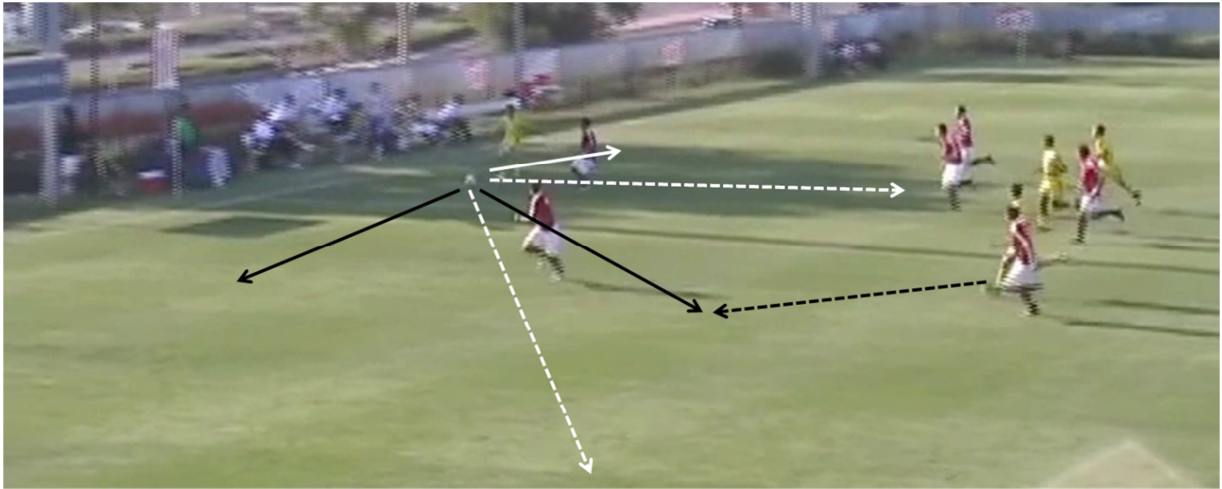


Figure 1: An example of a critical decision moment in a video trial. Dark options were deemed task-relevant by an expert panel. Light options are examples of other options participants may generate, but were deemed task-irrelevant.

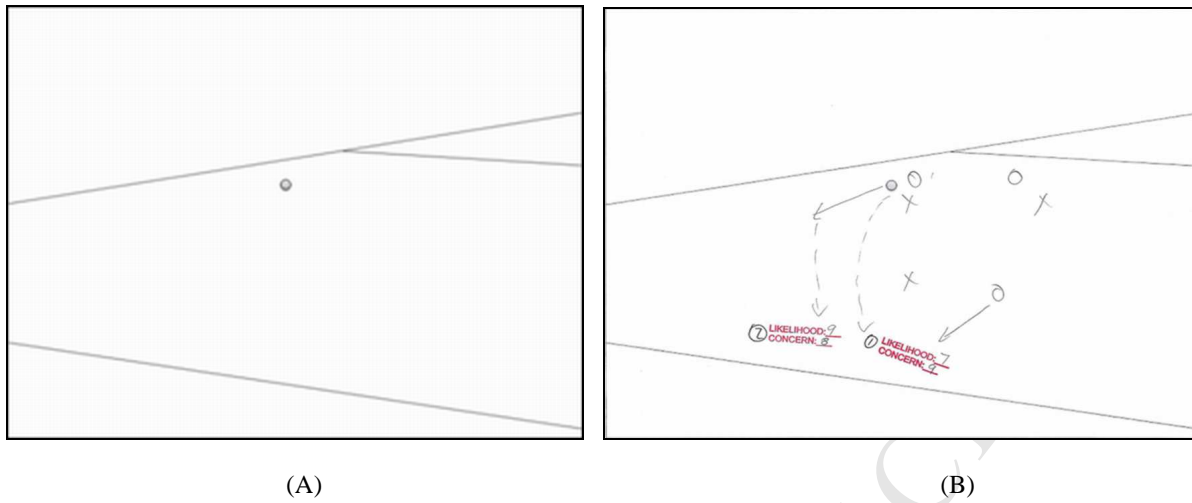


Figure 2: The occlusion image (A) and a sample response sheet (B) for the same trial. *Note.* A solid arrow denoted the player movement, a dashed arrow denoted ball movement. X and O were used to mark the position of defensive and offensive players, respectively.

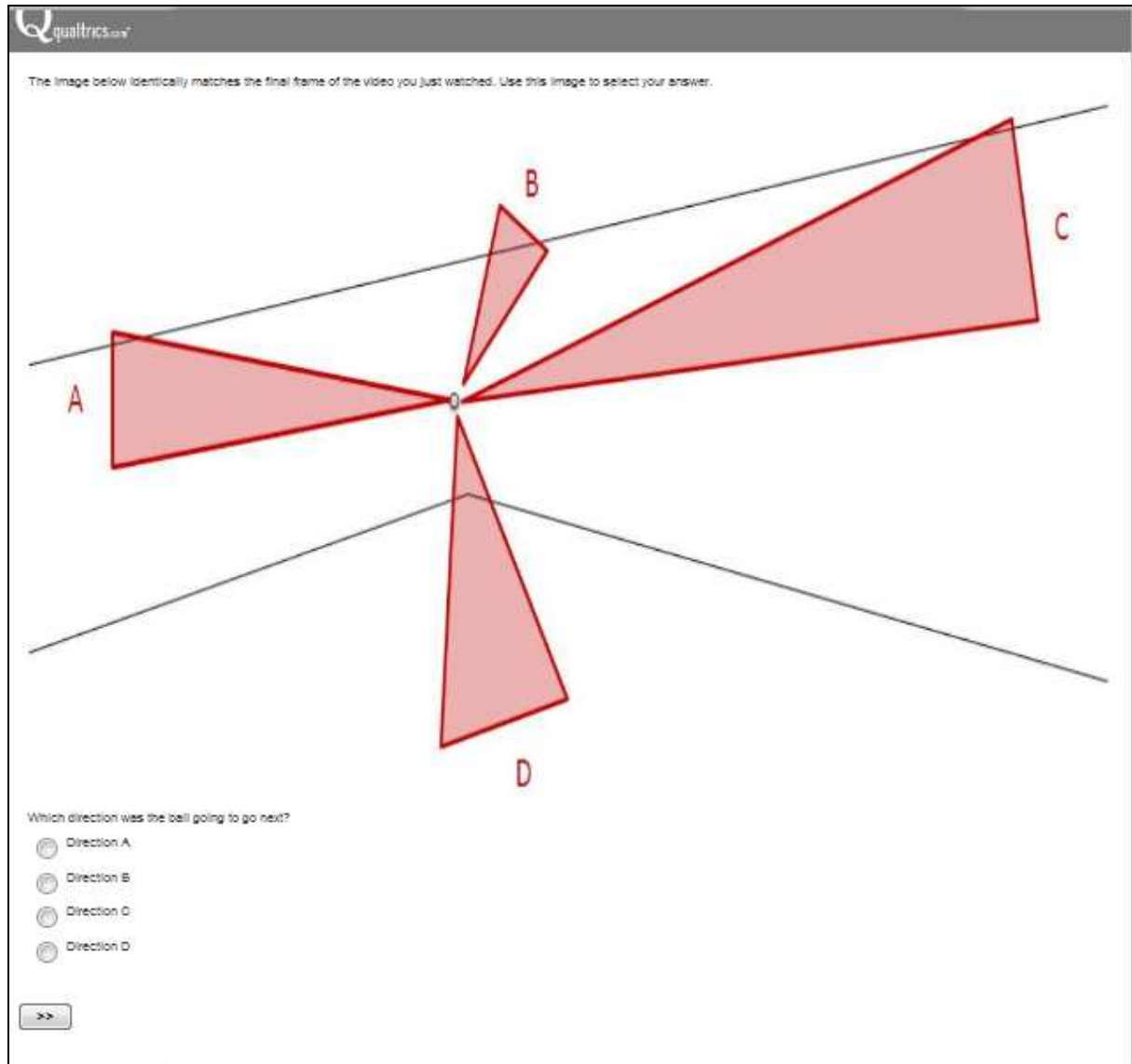


Figure 3: An example occlusion image from a *Direction* trial of the OASSIS.

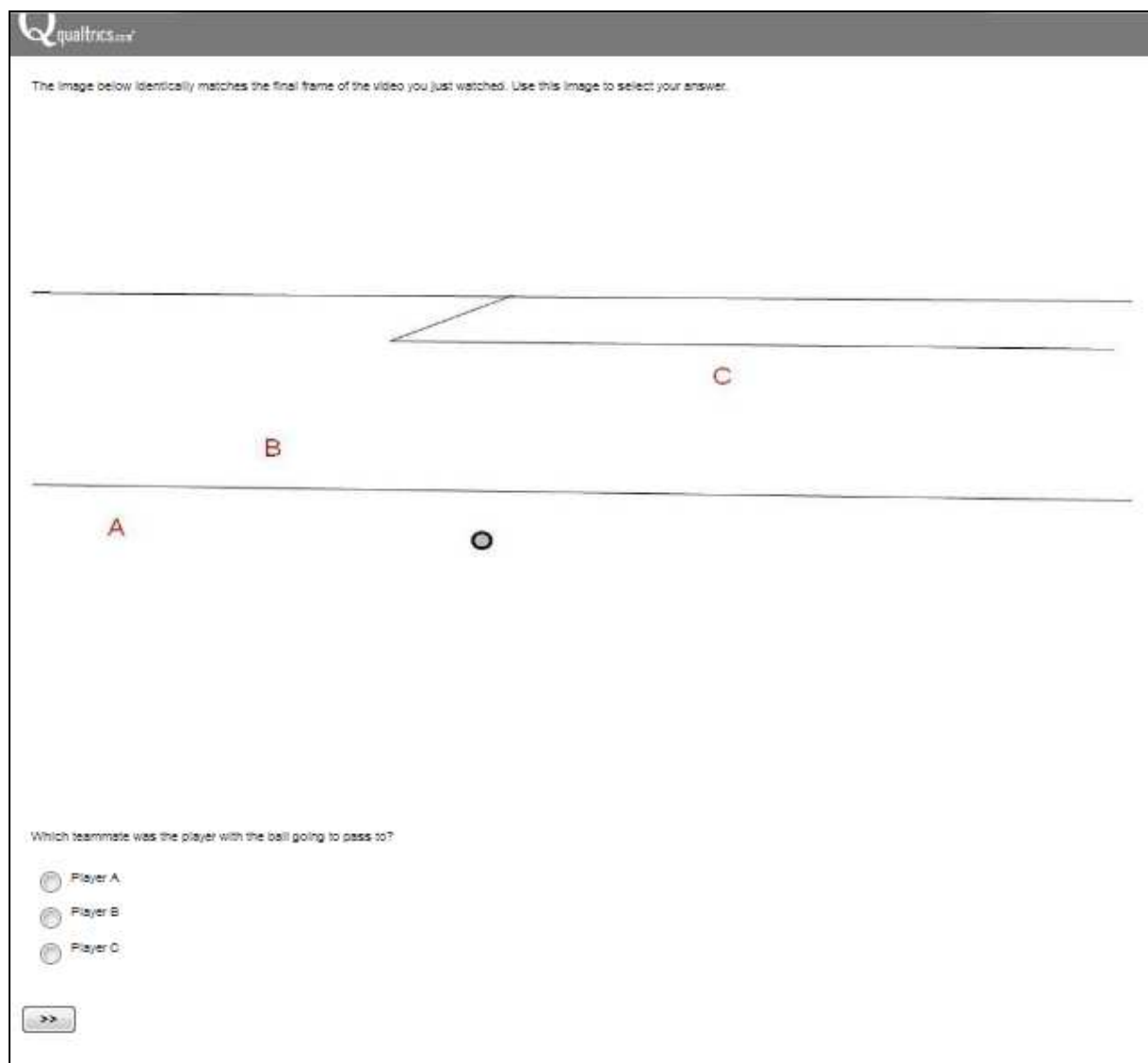


Figure 4: An example occlusion image from a *Pass Recipient* trial of the OASSIS.

Highlights

- We provide support for LTWM theory in soccer performance.
- We evaluate LTWM theory and the TTF heuristic.
- We introduce an online test of perceptual-cognitive skill in soccer.

ACCEPTED MANUSCRIPT

Appendix A: The Berlin Numeracy Test (BNT; see Cokely et al., 2012)

- 1) Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? Please indicate the probability in percent.

____%

- 2) Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3, or 5)?

____ out of 50 times.

- 3) Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws how many times would the die show the number 6?

____ out of 70 throws.

- 4) In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom is red? Please indicate the probability in percent.

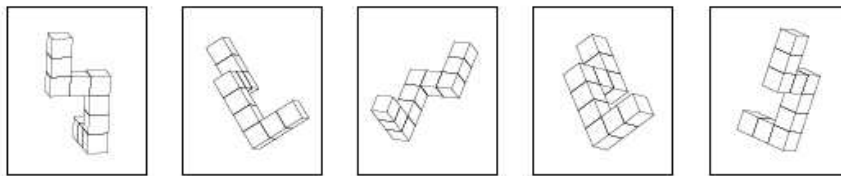
____%

MENTAL ROTATIONS TEST (MRT-A)

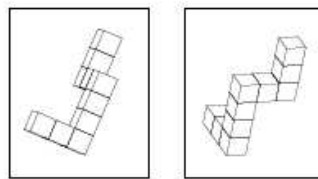
This test is composed of the figures provided by Shepard and Metzler (1978), and is, essentially, an Autocad-redrawn version of the Vandenberg & Kuse MRT test.

©Michael Peters, PhD, July 1995

Please look at these five figures



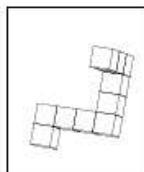
Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object (or yourself with respect to the object), as you look from one drawing to the next.



Here are two drawings of a new figure that is different from the one shown in the first 5 drawings. Satisfy yourself that these two drawings show an object that is different and cannot be "rotated" to be identical with the object shown in the first five drawings.

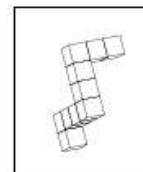
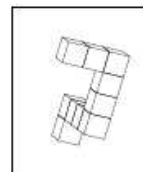
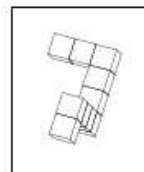
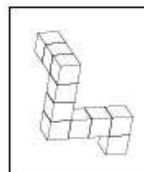
Now look at this object:

1.



Two of these four drawings show the same object.

Can you find those two? Put a big X across them.



If you marked the first and third drawings, you made the correct choice.