Please cite the Published Version

Metacognitive judgements of perceptual-motor steering performance

Mole, C. D.[1], Jersakova, R.[1], Kountouriotis, G. K.[2], Moulin, C. J. A.[3] & Wilkie, R. M[1]*

[1]*School of Psychology, University of Leeds, Leeds, United Kingdom*

[2]*Department of Psychology, Manchester Metropolitan University, Manchester, United Kingdom*

[3] *Laboratoire de Psychologie et Neurocognition (CNRS 5105), Université Grenoble Alpes, Grenoble, France*

* Corresponding author: Dr Richard Wilkie, School of Psychology. University of Leeds, Leeds, LS2 9JT. Email: R.M.WILKIE@leeds.ac.uk

Abstract

Control of skilled actions requires rapid information sampling and processing, which may largely be carried out subconsciously. However, individuals often need to make conscious strategic decisions that ideally would be based upon accurate knowledge of performance. Here we determined the extent to which individuals have explicit awareness of their steering performance (conceptualised as 'metacognition'). Participants steered in a virtual environment along a bending road whilst attempting to keep within a central demarcated target zone. Task demands were altered by manipulating locomotor speed (fast/slow) and the target zone (narrow/wide). All participants received continuous visual feedback about position in zone, and one sub-group were given additional auditory warnings when exiting/entering the zone. At the end of each trial participants made a metacognitive evaluation: the proportion of the trial they believed was spent in the zone. Overall, whilst evaluations broadly shifted in line with task demands, participants showed limited calibration to performance. Regression analysis showed that evaluations were influenced by two components: i) direct monitoring of performance, and ii) indirect task heuristics estimating performance based on salient cues (e.g. speed). Evaluations often weighted indirect task heuristics inappropriately, but the additional auditory feedback improved evaluations seemingly by reducing this weighting. These results have important implications for all motor tasks where conscious cognitive control can be used to influence action selection.

Introduction

Locomotor control is a ubiquitous human behaviour that is essential for successful interactions with the world, moving through space to achieve a goal. Whilst many of the underlying locomotor behaviours that support slow-paced walking also 'scale up' to faster modes of transport (e.g. cycling), when driving a vehicle on a road there are a host of new behaviours that need to be learnt in order to be successful (changing gears, navigating at high speeds, interacting with vulnerable road users, monitoring road rules, using mirrors etc.). Consequently, driving is one of the most risky perceptual-motor tasks that humans carry out routinely, yet there is limited understanding of the cognitive mechanisms that support this task.

Driving can been conceptualised as a hierarchical task (cf. Michon, 1985; McRuer & Allen, 1977), where the execution of the 'low-level' perceptual-motor behaviours (that move the human along the desired trajectory) are supervised by guidance systems that are responsible for making 'higher-level' trajectory planning decisions (such as when to overtake or how tightly to cut a corner). The basic principles underlying steering control are fairly well established, with the human central-nervous system coordinating fast perceptual-motor feedback loops to correct immediate errors, combined with the use of prospective 'future path' information to guide steering 1-2s ahead (Land & Lee, 1994; Land & Horwood, 1995; Salvucci & Gray, 2004; Wilkie et al., 2008; Frissen & Mars, 2014; Mole et al., 2016; Lappi, 2014). One of the critical properties of successful steering control is the rapid production of appropriate actions. For this reason, the interactions between steering responses and slower 'cognitive' processes (e.g. memory) are likely to be limited. When driving, however, locomotor scenarios tend to be more complex than simply steering a bend, and there are cognitively penetrable strategic decisions influencing the initiation (e.g. deciding to pull out at a junction) or maintenance (e.g. braking or steering to avoid a crash) of certain actions. Such decision-making requires accurate self-assessment of competence to avoid an individual placing themselves in a situation where the task demands exceed their capabilities (Sundström, 2011).

The extent to which drivers have accurate knowledge of their perceptual-motor performance is currently unclear. In terms of crude self-assessments of skill level, the driving population is notorious for overestimation, with the majority considering themselves "better-than-average" drivers (Svenson, 1981; Goszcynska & Roslan, 1989; see Kovácsová et al., 2016 for a replication of this effect in cyclists). Such inflated self-assessment has been linked to an increase in accidents in novice drivers (Gregersen, 1996), as well as to ratings of unsafe driving behaviour (Freund, Colgrove, Burke, & Mcleod, 2005; Amado, Arikan, Kaça, Koyuncu, & Turkan, 2014). It has been pointed out, however, that asking drivers to estimate their general driving skill may not actually assess their 'awareness of performance' (Sundström, 2008) since above-average social

comparisons appears to apply to many human judgements of ability (Dunning, Heath, & Suls, 2004; Sharot, 2011; Simons, 2013), especially when information about the trait or aspect of skill being compared is uncertain (Dunning et al., 1989). Instead, it has been proposed that awareness of driver performance should be assessed by asking participants to make judgements of performance on a quantifiable driving task (Sundstrom, 2008), and there is some evidence that self-evaluations can be improved if the evaluation criterion is specific (Amado et al., 2014). Consequently, there have been attempts in the Traffic Psychology literature to examine driver *calibration* – the degree of alignment between subjective assessment of performance and objective performance measures (Horrey, Lesch, Mitsopoulos-Rubens, & Lee, 2015). When the driving task is quantifiable (such as quantity of traffic cones hit), Roberts et al. (2016) found that ratings of performance can be fairly well-calibrated to actual performance.

These methods are useful to assess general self-perceptions of performing well (or not so well), which may feed into conscious preparatory decisions regarding driving risk and capability, for example deciding whether to drive at night (Fuller, 2005). Moment-to-moment trajectory planning decisions, however, presumably require precise information about 'on-line' perceptual-motor performance in order to appropriately guide steering control (McRuer et al., 1977). However, the extent to which humans can accurately monitor perceptual-motor performance is unclear. Mamassian (2008) demonstrated that people have limited knowledge about motor performance variability, leading to sub-optimal, overconfident, motor decisions in a timing task. There is also evidence that some aspects of perceptual-motor behaviour are largely automatic (and unconscious) with individuals seemingly able to correct errors without being explicitly aware of making the corrections or even that corrections were needed (Fourneret & Jeannerod, 1998; Sinanaj, Cojan, & Vuilleumier, 2015; Prablanc & Martin, 1992; Pelisson et al., 1986).

A useful framework for examining awareness of performance is metacognition. The ability to successfully monitor cognitive processes has been demonstrated in various domains, from memory (see Dunlosky & Bjork, 2008), to (visual) perceptual decisions (e.g. Fleming et al., 2015; Peters & Lau, 2015) and action-control monitoring (i.e. sense of agency or feeling in control of one's actions, e.g. Sidarus & Haggard, 2016; Metcalfe, Eich & Miele, 2013). Metacognition is linked to the control of behaviour (Nelson & Narens, 1990) in that metacognitive monitoring has behavioural consequences. For example, when participants make metacognitive judgments about how well they have learned new information (word pairs), this has direct influence on what items they choose to restudy before a final memory test, even in situations where these judgments do not accurately reflect how well the information has been learned (Metcalfe & Finn, 2008). Understanding how monitoring judgments are constructed is therefore one of the key questions in the metacognitive literature and it is particularly relevant for judgments made

during high-speed locomotion, where there is greater risk of poor judgements leading to harmful errors (Freund et al., 2005).

The prevailing framework for capturing metacognitive function stipulates two main levels (Nelson & Narens, 1990): an object-level (task performance) which is monitored and controlled by a meta-level (performance awareness). This two-level metacognition framework (Nelson & Narens, 1990) has pleasing parallels with the guiding level and execution level proposed in driving (McRuer et al., 1977), but to date there has been limited research into metacognitive awareness of perceptual-motor behaviour (but see Mamassian, 2008). In the domain of motor learning (where novices are taught new motor skills) it has been shown that errors in monitoring can occur, with individuals exhibiting flawed knowledge of how well they have learnt a motor skill (Simon & Bjork, 2001; Abushanab & Bishare, 2013), or lacking insight into the optimum strategy with which to learn that skill (Simon & Bjork, 2002; see Soderstrom & Bjork, 2015 for a review). Once the skill is learnt, motor performance will still require monitoring to ensure changes in performance are not invalidating strategic level decisions. It remains to be demonstrated whether drivers have good metacognitive awareness of perceptual-motor steering performance.

Metacognition is primarily viewed as an inferential process (Koriat, 2000) which relies on *cues* derived from the task at hand to construct judgments about performance. An example of a task related cue that has been shown to influence metacognitive judgments in a range of tasks is processing fluency of the monitored stimulus (e.g. Rhodes & Castel, 2008; Sidarus & Haggard, 2016). While the cues that influence metacognitive judgments can be related to task performance (e.g. Koriat, 2008), in some instances participants fail to act on performance-relevant information (Koriat & Bjork, 2005) or, alternatively, they base their monitoring judgments on irrelevant information that does not relate to task performance (e.g. Rhodes & Castel, 2008).

A similar cue-based approach has been proposed as a conceptual framework for assessing driver calibration (Horrey et al., 2015). Good calibration arises from selecting perceptual cues that capture task performance, whereas miscalibration can result from selecting inappropriate cues or failing to weight cues appropriately (Horrey et al., 2015). However, given the evidence that specific perceptual-motor tasks are executed with limited cognitive penetrability (Fourneret & Jeannerod, 1998; Sinanaj, Cojan, & Vuilleumier, 2015; Prablanc & Martin, 1992; Pelisson et al., 1986), it remains unclear whether

drivers are able to select and appropriately weight perceptual cues when making judgements of performance. Furthermore, Roberts et al. (2016) found that the provision of feedback did little to improve calibration, suggesting that drivers may have difficulty using additional feedback to re-weight cues. The literature that examines judgements of performance in driving has tended to use real-world settings (e.g. Amado et al., 2014; Roberts et al., 2016). However, in the real-world it is difficult to control the cues available to drivers, or even to measure the feedback that was available whilst moving through the scene. This makes it difficult to determine how judgements were generated, and which sources of information judgements were actually based upon.

To attempt to address these issues we used a computer simulated virtual driving environment in order to examine metacognitive judgements using a highly controlled steering task where the parameters influencing performance are specific and explicit, allowing direct comparisons between judgements and performance across different task conditions. The current experiment sets out to investigate i) whether individuals could accurately monitor performance on a specific perceptual-motor task across conditions of varying difficulty, ii) which task-relevant cues informed judgements, iii) whether judgements are altered with available feedback.

Individuals were required to steer along a bending road and were asked to keep within a sub-region - a central, delineated, target zone (which was narrower than the road itself). Two task parameters were varied: speed of travel, and the width of the target zone. Faster speeds and narrower zones effectively increased task difficulty (and so should impair performance), ensuring that there was variation in performance across conditions. After each trial participants made a judgement of what proportion of the trial they spent within the required zone, and then determined how well judgements reflected the performance.

In order to investigate whether quality of feedback could improve judgments, we split participants into two groups. The first group received the standard continuous visual feedback provided by the visible scene (i.e. position in lane). The second group received the same visual feedback, as well as additional auditory feedback in the form of a short auditory tone whenever they entered or exited the target zone. The auditory feedback was temporally compatible with the near portion of the zone, which can be used for monitoring lateral road position (Salvucci & Gray, 2004; Mole et al., 2016), therefore should serve to increase the salience of task-relevant road position[1].

---

[1] Post-trial knowledge of results could have been an alternative method to provide feedback, but it seems to have mixed effects on performance monitoring (Carter, Smith, & Ste-Marie, 2016; Roberts et al., 2016) and is a relatively artificial way of providing feedback for trials of this duration (i.e. drivers do not routinely experience explicit knowledge of results in the real-world except for

the occasional major error such as a crash/collision). In contrast, drivers routinely incorporate multi-sensory signals into lane-keeping (such as the haptic feel of the steering wheel, or the sound caused by the vibration of the tyres passing across lane markings or rumble strips).

Method

Participants

64 participants (34 females, 30 males; aged 17 to 52, mean = 24.95 yrs, std = 7.74) took part in the study, split into two groups of 32. All participants had normal or corrected-to-normal vision. 58 participants had full driving licenses (average number of years since test = 6.94 yrs) and the remaining six participants had relevant experience (in the form of driving lessons or go-karting) All participants gave informed consent and the study was approved by the University of Leeds Ethical Committee, and complied with all guidelines as set out in the declaration of Helsinki.

Apparatus

The experiment used a PC (Intel i7 950 3.07 GHz) with an update rate of 60Hz. The stimuli were programmed using WorldViz Vizard 3.0 and were back-projected (PLC-XU58) onto a 1.98m x 1.43m screen (as per Mole et al., 2016). Participants were sat in a darkened matt-black viewing booth, in a height-adjustable racing-style driving seat with an eye position 1m away from the screen and 1.2m from the ground, so the total visual angle of the display was 89.42° x 71.31°. A force-feedback wheel (Logitech G27) was used to control steering.

Stimuli

The visual stimulus consisted of a tessellated gravel texture applied to the ground-plane, with two superimposed white lines that acted as road edges, and a road sub-region marked by two green lines that acted as the target steering zone (see Figure 1A). Similar to previous studies (Kountouriotis, Floyd, Gardner, Merat, & Wilkie, 2012; Mole et al., 2016) the curvature and width of the road (white lines) were kept constant throughout conditions (3m wide, 60m radius). The first 9m of each road consisted of a straight part followed by a constant curvature bend. The driver started in the centre of the road in every trial. In half the trials the target zone (green lines) was 0.3m wide (10% of road width; "Narrow" conditions), in the other half the trials were 0.5m wide (16.7% of road width; "Wide" conditions). The speed of locomotion was constant within each trial; in half of the trials it was set at 12 m/s (26.84mph; "Slow" conditions) and in half of the trials it was set at 18 m/s (40.26mph; "Fast" conditions).

The difference between participant groups was the presence/absence of auditory feedback linked to road position. One group received just *visual feedback* (VF) and no additional auditory feedback. The second group received both *auditory and visual feedback* (AVF) - whenever the driver crossed the green lines denoting the target zone a short beep [240Hz tone] was sounded for .1s. Thus these participants received auditory feedback whenever they entered or exited the zone.

When a typical adult steers round a bend, eye-movements are directed towards the future path, with near-road information mostly sampled via peripheral vision (Wilkie et al., 2008). The monitoring task (described below) required near-road positional information, we therefore anticipated that participants would direct gaze to the near portion of the road in order to disambiguate road position signals normally picked up through peripheral vision (which has reduced acuity). On the other hand, the AVF group have road position signals augmented by an auditory beep so may not feel the need to redirect gaze to the near-road. Such vertical relocation of gaze behaviour lower in the scene can influence the steering trajectories (Mole et al., 2016) and may alter the nature of the task from a 'steering-a-bend' task (using visual preview) to a 'error nulling' task. In order to keep the steering task similar to driving a vehicle, and avoid any possible gaze differences between groups, we controlled eye-movements by spatially placing a red fixation cross 16m ahead of the driver in the centre of the road. Participants were asked to fixate on the cross, and were reminded during the experiment to continue fixating on the cross. This method has been used in a number of eye-tracking studies and participants are able to adequately fixate on the cross (e.g. Kountouriotis et al., 2012; Mole et al., 2016).
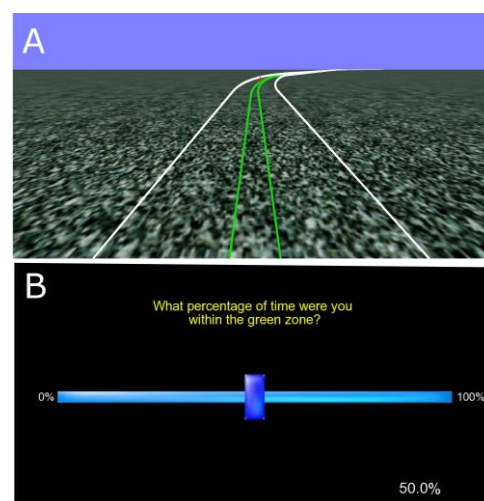
Figure 1. A) Snapshot of the experimental stimulus showing the Narrow (0.3m) condition. A 3m road was always visible (white outline) but participants were instructed to stay within the target zone (green outline). Participants were asked to fixate on the red cross, which was placed 16m in front of the participant at the middle of the road. **B)** The post-trial judgement screen. Participants used the wheel to manipulate the slider and gear pads to record judgements and initiate the next trial.

At the end of each trial participants were presented with a sliding scale from 0-100% with the question "What percentage of time were you within the green zone?" (Figure 1B). Participants could manipulate the slider with the steering wheel and see what percentage they were selecting by an updating text label. Once the participant had reached a decision they could select the score by clicking the steering wheel gear-pads. To avoid any bias linked to using the wheel to determine the value, the steering wheel rotation was either translated directly onto the scale, or reversed (i.e. turning the wheel right would move the slider left). The mapping was randomised trial-to-trial to avoid any biomechanical priming effects (e.g. on right bends the participant may have been more willing to make a higher percentage judgements as they are already turning the wheel in the direction of 100%) since motor movements are known to influence preferences (Beilock & Holt, 2007). Whilst it is possible that any priming effect would have been 'washed out' since half the bends were leftwards and half the bends were rightwards, an added reason for mirroring the wheel was to try to evoke carefully considered judgements.

Procedure

Participants were given five practice trials to ensure they knew what was required (to fixate the cross and steer within the target zone) and familiarise themselves with the experimental setup and steering wheel dynamics. To avoid any confusion as to what constituted '"keeping in the zone" participants were asked to keep their midline within the zone. After each trial the participant would make a percentage judgement of total trial time spent within the target zone (see Figure 1B) then initiate the next trial. Participants were asked to return the wheel back to the centre immediately after recording a judgement to avoid starting the next trial with the steering wheel at a large angle. The participant was asked to fixate on the cross, and reminded to fixate once during the experiment.

The experiment had two within subject factors (Speed and Width) and one between subjects factor (Level of Feedback). This resulted in four conditions per group, with each group completing 24 trials (six repetitions of each condition, 3 right bends and 3 left bends). The order of trials was randomised for each participant with no more than three consecutive trials occurring of the same condition. Each trial lasted seven seconds (excluding the time taken for the participant to make their judgement, which was not limited), resulting in a minimum experiment running time of 2 minutes and 48 seconds.

Analysis

Road position was recorded frame-by-frame. The number of frames the participant spent outside the target zone was calculated, which was converted to a Percentage in Zone (PiZ) score to enable direct comparison against the participant judgements. Analysis included the whole trial because the participants' judgements were similarly based on the whole trial. The second key measure was the participant's post-trial Judgement of Performance (JoP): the percentage of total trial time the participant thought they had spent within the confines of the central target zone.

We assessed the ability of participants to judge their performance through three metacognitive measures: sensitivity, calibration, and resolution. First we determined whether judgements were sensitive to changing parameters of the task by examining whether JoP scores vary according to level of feedback and task difficulty (*metacognitive sensitivity*). We then compare JoPs directly to PiZ scores by taking the signed and absolute difference scores (signed & absolute accuracy), to assess the extent that judgements were calibrated to performance (*metacognitive calibration*). However, in this experiment competent monitoring of perceptual-motor actions may not require that the participant's internal judgement scale is perfectly matched with the external performance scale i.e. participants do not need to be able to precisely calculate how many frames of each trial they spent within the zone as long as their judgements reliably track fluctuations in performance. Indeed, a hesitant individual could have a propensity to consistently underestimate performance therefore score badly on metacognitive *calibration* measures, but their judgements may perfectly capture subtle changes in task performance therefore showing excellent *resolution.* Therefore, we also calculate *metacognitive resolution* by using Pearson's R (Schraw, 1995) to compute intra-individual correlations between JoP and PiZ. All measures were subject to a 2 (Speed) x 2 (Width) x 2 (Level of Feedback) ANOVA using R.

Results

Performance (PiZ)

Participants attempted to steer within the target zone, and task performance was measured using the percentage of time within the zone. Table 1 displays the mixed model ANOVA results: there were main effects of Zone Width, Speed, and Feedback. There were also significant interactions between Width x Speed and Width x Speed x Feedback (Figure 2A). As expected, for both feedback groups participants kept within the target zone for a longer proportion of time on wider and slower trials, with the additional feedback increasing the amount of time spent in the zone. The interactions occurred because additional auditory feedback led to a greater improvement between widths when travelling at fast speeds compared to slow speeds (shown by a steeper slope for the dashed line with black symbols compared to the solid line with black symbols in Figure 2A). Though group performance did not reach 100% for any conditions, performance in the Wide-Slow (AVF) condition was >90% so ceiling effects in

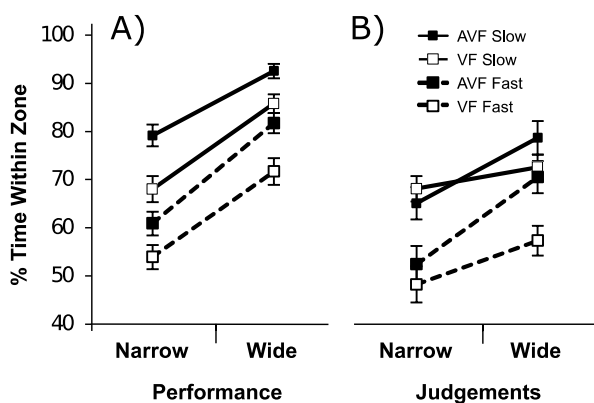some individuals may have artificially reduced the effect of width.

Figure 2. Average PiZ (A) and JoP (B) scores for participants with visual feedback only (empty symbols) or additional auditory feedback (black symbols) at fast speeds (dashed lines) or slow speeds (solid lines) with Narrow or Wide zones. Bars represent standard error of the mean.

Metacognitive Sensitivity (JoP)

The steering performance measures showed that task conditions altered the ability of participants to stay in zone. At first pass we wished to see whether JoPs were sensitive to changes in task difficulty and provision of additional feedback. The results of the mixed model ANOVA are displayed in Table 1. There were main effects of Speed and Zone Width, indicating some sensitivity to these conditions

Interestingly, there was not a main effect of Feedback indicating that there was not a consistent shift in JoP scores when auditory feedback was available. There were, however, significant interactions between Width x Feedback, Speed x Feedback, and Width x Speed (shown in Figure 2B). The interactions showed that: auditory feedback had a greater effect on JoPs when the zone was wide; Speed had a weaker effect for the AVF group; and for both groups, the magnitude of JoPs increased more between widths when travelling at fast speeds compared to slow speeds.

Table 1. The effect of zone width, locomotor speed and levels of feedback on PiZ and JoP scores. Produced by mixed design 2 (Speed) x 2 (Width) x 2 (Level of Feedback) ANOVAs, with Eta squared effect sizes reported. Degrees of freedom was 1, 62 throughout.

	Performance (PiZ)			Judgements (JoP)		
	F	η^2	P	F	η^2	P
Width	438.12	.26	<.001*	63.49	.079	<.001*
Speed	171.21	.19	<.001*	92.58	.14	<.001*
Feedback[F]	5.8	.07	.019*	2.41	.029	.125
Width x Feedback	.08	<.001	.78	9.64	.013	.003*
Speed x Feedback	.25	<.001	.62	7.85	.014	.007*
Width x Speed	5.22	.004	.026*	6.21	.003	.015*
Width x Speed x Feedback	4.45	.003	.039*	.089	<.001	.767

[F]Level of Feedback was the only between-subjects factor

*p<.05

.

Metacognitive Calibration (Signed & Absolute Accuracy)

In order to assess *Metacognitive Calibration* we directly compared JoP scores with PiZ scores. Whilst signed accuracy (Figure 3A) gives a useful indication of the overall tendency to over/under estimate, if judgments are inconsistent and include both over/under estimations then this could average to near zero and fail to capture the variability. In contrast, absolute accuracy (Figure 3B) gives an indication of total precision error, regardless of error direction. We therefore calculated both Signed and Absolute Accuracy measures across each condition.

The results of the ANOVA are displayed in Table 2. For Signed Accuracy (Figure 3A), there was a significant main effect of Zone Width, but not a main effect of Speed or Feedback. There was, however, significant Width x Feedback and Speed x Feedback interactions (it is worth noting, however, that the effect sizes for both the main effect of Width and the two interactions are small). As shown in Figure 3A, in VF conditions there was a large tendency to underestimate in Wide compared to Narrow conditions (where underestimation was reduced or eliminated). In contrast, for the AVF group both Narrow and Wide conditions produced a tendency to underestimate (the Width x Feedback interaction). The Speed x Feedback interaction occurs because Fast speeds produced greater underestimating behaviour in VF, but the effect was reversed in AVF (Slow speeds produced greater underestimating behaviour).

For Absolute Accuracy (Figure 3B), the ANOVA revealed only a marginally significant main effect of Width, with a small effect size (Table 2). This was due to JoPs having greater accuracy on Wide Zones compared to Narrow Zones (Figure 3B). As can be seen in Figure 3B, this effect was mostly due to increased accuracy for the Wide Zones during AVF.

The magnitudes for signed accuracy (Figure 3A) are less than absolute accuracy (Figure 3B), showing that the tendency to under/over estimate was not completely consistent across participants (identical magnitudes would indicate that the direction of under/over estimation was the same for every participant). Taking both accuracy measures together, it seems that although the average tendency to over/under estimate varied between conditions (Figure 3A), the relative over/under estimation precision error was fairly consistent and also large (around 20 percentage points; Figure 3B) – suggesting that participants were poorly calibrated.
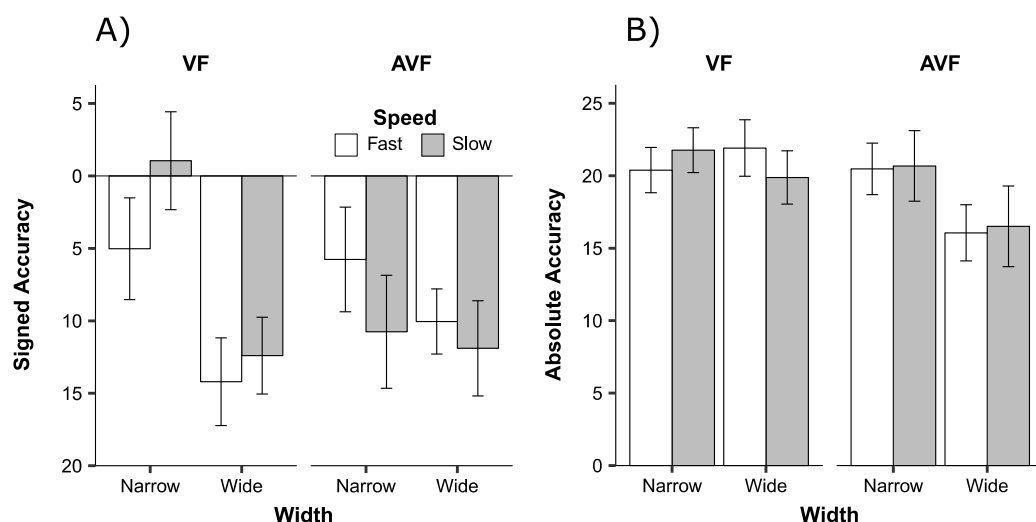

Figure 3. Metacognitive Calibration, captured by the (A) Signed, and (B) Absolute trial-by-trial differences between judgement and performance scores. In A) positive scores indicate overconfidence. VF= Visual feedback only; AVF= Auditory and Visual feedback. Error bars represent standard error of the mean.

Table 2. The effect of zone width, locomotor speed and levels of feedback on Metacognitive Calibration and Metacognitive Resolution. Produced by mixed design 2 (Speed) x 2 (Width) x 2 (Level of Feedback) ANOVAs, with Eta squared effect sizes reported. Degrees of freedom was (1, 62) throughout.

	Calibration: Signed Accuracy			Calibration: Absolute Accuracy			Resolution (Pearson's R)		
	F	η^2	P	F	η^2	p	F	η^2	p
Zone Width	26.07	.036	<.001*	4.07	.01	.048*	.7	.002	.41
Speed	.032	<.001	.86	<.001	<.001	.99	6.92	.026	.011*
Feedback[F]	.24	.003	.63	1.19	.013	.28	.14	.001	.71
Width x Feedback	9.79	.014	.003*	3.45	.008	.07	.009	<.001	.93
Speed x Feedback	6.64	.01	.012*	.11	<.001	.74	3.17	.012	.08
Width x Speed	.085	<.001	.77	1.32	.001	.25	1.01	.003	.31
Width x Speed x Feedback	3.71	.003	.058	1.76	.002	.19	3.31	.009	.074

[F]Level of Feedback was the only between-subjects factor
*$p<.05$.

Metacognitive Resolution (Pearson's R)

The calibration measures suggest that participants were unable to closely match judgements to performance. However, this does not necessarily capture whether participants were sensitive to changes in performance (i.e. metacognitive resolution). Pearson's R values were computed across PiZ and JoP scores for each individual within each group for each condition. The resulting R values indicate the extent to which individual judgements reflected the performance in those conditions, and provide a measure of relative accuracy. The pattern across conditions/groups was analysed using a 2 (Speed) x 2 (Width) x 2 (Level of Feedback) ANOVA (see Table 2) which revealed a main effect of Speed (with a small effect size), but not a main effect of Width or Feedback, nor any interactions. The main effect of speed was due to significantly higher metacognitive resolution when travelling faster at 18m/s (grand mean = .5) compared to travelling slower at 12m/s (grand mean=.36; Figure 4). Whilst in some conditions there were isolated differences in resolution between levels of feedback (e.g. for Slow-Narrow providing auditory feedback increased resolution from 25% to 47%), the results of the ANOVA demonstrate

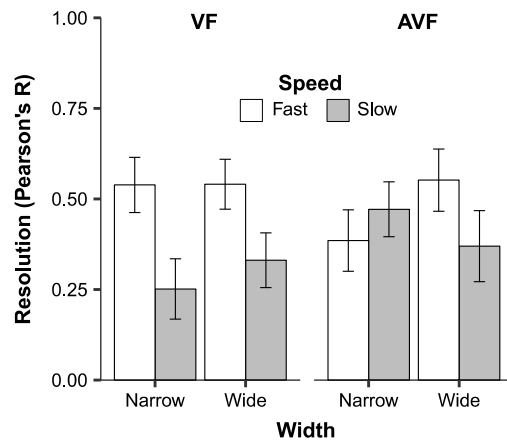that these differences did not reach significance.

Figure 4. Metacognitive Resolution for each condition and each level of feedback (VF= Visual feedback only; AVF= Auditory and Visual feedback). Error bars represent standard error of the mean.

Investigating Metacognitive Mechanisms

One aim of the study was to examine whether task-related cues were used to generate metacognitive judgements. The

metacognitive measures of *calibration* and *resolution* demonstrate that participants' judgements seem to be poorly calibrated (and mostly characterised by under-estimation of time in zone), with only limited resolution. The question remains *how* individuals were generating their judgements. Participants could have been attempting to directly monitor steering performance based on the measure of interest (PiZ), and that for some reason participants were simply unable to determine when their midline moved in or out of the zone (see Discussion for a consideration of some possible methodological reasons for this). On other hand, it could be that in attempting to gauge performance, participants' were (consciously or unconsciously) using task-relevant cues as heuristics, such as speed of travel or lane width (e.g. Koriat, 2008). Such a strategy could lead to both poor calibration and resolution if participants do not weight the cues appropriately.

In order to examine this issue further we carried out a trial-level regression analysis, following a method proposed by Lorch and Myers (1990; see also Allen & Hulme, 2006; Metcalfe et al., 2013). Within this analysis, each participant's JoPs were submitted to a separate within-subjects regression analysis in which Performance (PiZ for that trial), Speed and Zone Width (represented by binary flags: 1 = Slow, 0 = Fast; 1 = Wide, 0 = Narrow) were entered together as predictors. The resulting beta values for each predictor were then analysed using a one-sample *t*-test to examine whether they were significantly different from 0. This allowed us to establish whether each predictor contributed to JoPs, whilst controlling for inter-participant variability. Results for the three predictors of interest for both Feedback groups are presented in Table 3.

The regression analysis demonstrates that actual task performance and task-relevant cues each contributed to determining judgements of performance (Table 3). For VF conditions, participant judgements were determined mostly by performance, and to a lesser extent a Speed heuristic that altered judgements when trials were fast or slow (JoPs increased when speed slowed). Whilst the beta value for a Zone Width heuristic indicates that JoP scores, on average, actually *decreased* when the zone became wider (after taking into account task Performance and a Speed heuristic), this trend was variable between participants (shown by a large beta standard error) so was not significant (Table 3). When auditory feedback was added (AVF), judgements were similarly determined mostly by performance, but the heuristics used to influence judgements were based upon both Speed and Zone Width. Although the beta value *magnitudes* of both cues are less than those of VF, a Zone Width heuristic (higher JoPs for wider zones) appears to consistently contribute to JoPs across participants. Speed cues appear to have a diminished role in AVF (compared to VF) but still contributes significantly to judgements.

The predictive weight of actual task performance remained fairly consistent across feedback conditions - this is in line with the results for Metacognitive Resolution, which did not reveal significant differences between feedback conditions when predicting judgments using PiZ scores only (see Figure 4)

Table 3. Mean betas for each variable included in the within-subjects regression analyses predicting JoP by Feedback group. Results of the one-sample t-tests evaluating whether the mean beta values were significantly different from 0 are also reported. For Width and Speed, positive beta values indicate that JoP scores increased as the task became slower or the zone became wider.

Feedback	Predictor	β	SE β	t	df	P
VF	Zone Width	-.209	.134	-1.56	31	.129
	Speed	.264	.040	6.65	31	<.001*
	Performance	.504	.052	9.68	31	<.001*
AVF	Zone Width	.107	.038	2.77	31	.009*
	Speed	.074	.032	2.35	31	.025*
	Performance	.528	.075	7.08	31	<.001*

*$p<.05$.

Discussion

This study examined the extent to which participants were sensitive to changes in steering performance across conditions of varying difficulty. Whilst participant judgements did broadly reflect changes in performance, they were not particularly well-calibrated, and appear to be fairly coarse estimates that were partially based on salient task properties (e.g. fast vs. slow travel speeds). The key finding was that judgements of performance on a perceptual-motor task were generated using two interacting processes: i) explicit monitoring of performance, and ii) estimation of performance based on task heuristics (i.e. speed of travel or width of zone), the weighting of which was influenced by quality of feedback.

Participants were asked to steer within the confines of a central target zone and then make a retrospective judgement on their performance. Altering task parameters (Speed and Zone Width) and feedback availability produced systematic changes in task performance: faster and narrower conditions proved more difficult to steer than slower and wider conditions, and auditory feedback produced better steering performance in all conditions. The ability to incorporate auditory feedback into what is predominantly a visually guided action (e.g. Robertshaw & Wilkie, 2008; Wilkie, Kountouriotis, Merat, & Wann, 2010) highlights the robust nature of the sensorimotor system to use a flexible weighted combination of available cues (Ernst & Banks, 2002; Wilkie & Wann, 2002).

When examining how well judgements reflected these changes in performance we calculated three measures of success: sensitivity, calibration and resolution. Judgements appeared broadly sensitive to performance variations, estimating better performance for slower and wider trials, and when auditory feedback was present. This level of sensitivity fits well with evidence that people are somewhat sensitive (at least at a coarse level) to how changes in task parameters affect task difficulty (e.g. Roberts et al., 2016). However, judgements showed limited calibration when directly compared to trial performance, in most conditions consistently underestimating performance. This underestimation would seem to contradict the general observation across the driving literature of over-confidence in judgements of driving skill (Svenson, 1981; Goszcynska & Roslan, 1989). In the extant literature judgements are typically made with reference to an 'average' driver, rather than specifically to the actual performance on the task (as is the case here; see also Roberts et al., 2016). Here we have shown that self-judgements on a specific steering task can lead to performance underestimations rather than overestimations. However, previous research suggests that, when making motor planning decisions, people tend to overestimate motor precision (Mamassian, 2008). Similar overconfidence during our task would most likely produce inflated estimates of performance (since low steering variability would have led to increased time

within the zone). One possible cause for this mismatch is that there is a disconnect between motor decision making and self-report (i.e. they have different informational inputs), with the self-report underestimation being due to biased monitoring – for example isolated steering errors would have a small effect upon the actual PiZ score (since performance was calculated across the whole trial) whereas these errors may have been particularly salient to participants and thus they overestimated their impact upon PiZ. The degree to which cognitively penetrable self-report influences trajectory planning decisions is an important avenue for future research.

In the current study, the level of calibration was not the primary focus, since a participant may show poor calibration (e.g. due to response bias) whilst simultaneously perfectly adjusting judgement scores according to changes in performance across trials. However, values reported for monitoring resolution were fairly low (the grand mean R^2 for Monitoring Resolution was only .2), suggesting that in the steering task the driver's ability to pick up subtle fluctuations in performance was limited. It is possible that weak judgements were linked to methodological characteristics. For example, participants may not have been able to determine precisely where the 'midline' of the vehicle was (although whilst this may explain poor calibration, it does not explain limited resolution). It is also worth noting that there was no sign of fatigue or reduced engagement throughout the short experiment (when making judgements participants carefully selected their responses and took approximately 3.2s (SE=.12s) to do so).

It seems likely that the low monitoring resolution is symptomatic of the difficulties humans have trying to consciously monitor unfolding perceptual-motor events. This explanation is grounded in the literature that suggests that only some aspects of perceptual-motor behaviour may be cognitively penetrable (Prablanc & Martin, 1992; Pelisson et al., 1986; Pisella et al., 2000; Fourneret & Jeannerod, 1998; Sinanaj et al., 2015). Furthermore, the relatively poor monitoring performance shown here may well indicate that reports of well-calibrated driver self-assessments using fairly coarse measures (Sundstrom, 2011; Roberts et al., 2016) will have limited applicability to monitoring a precise perceptual-motor task during execution.

It is perhaps surprising that the group with auditory feedback did not have greater metacognitive resolution since this signal provided an explicit mark each time the driver entered or exited the zone. It does, however, echo Roberts et al., (2016) finding that feedback about driver performance does not clearly improve driver calibration. Roberts et al. (2016) suggested that in their study the delay between performance and feedback may have decreased the impact of feedback provision, but we have replicated this finding (at least with respect to monitoring resolution) when feedback is provided immediately in response to performance on the perceptual-motor task.

A major finding within the current study is that drivers may supplement limited performance monitoring

by using heuristic cue-based estimation. On a trial-by-trial regression analysis, we found that binary predictors of task characteristics significantly contributed to determining JoPs alongside actual performance scores (the exception being Zone Width in the VF condition). A cue-based approach to making self-assessments of driver performance has been recently proposed (Horrey et al., 2015), but we believe our findings support a potential generative mechanism of judgments of perceptual-motor performance that comprises two contributing processes: (i) explicitly monitoring actual performance and (ii) estimating performance based on task-relevant cues. These processes will be discussed in turn:

Explicitly monitoring actual performance

The regression analysis showed that task performance scores provided a (dominant) contribution to predicting JoPs, meaning that participants at some level monitor perceptual-motor performance related to steering behaviour and discriminate between this and the perceptual information relating to task parameters (i.e. Speed of travel and Width of Zone, which provided a smaller contribution to predicting JoPs). It remains unclear at what stage within the perceptual-motor system (i.e. motor planning, execution, or feedback) the monitoring process takes place. Fleming et al. (2015) identified a potential contribution (when making confidence judgements on a perceptual discrimination task) for premotor areas associated with motor planning but not areas associated with motor execution. This suggests that at least some of the contribution of PiZ to JoPs may be explained by motor planning (e.g. steering corrections). Additionally, it is possible that the monitoring limitations observed in our data may be at least partly due to perceptual limitations in detecting deviations of the midline relative to the target zone, rather than limitations in metacognitive monitoring ability per se (these two processes appear to have distinct associated brain areas, Sinanaj et al., 2015). It is worth noting, however, that Performance weighting changed little between feedback groups, suggesting that the 'monitoring of performance' aspect of judgement generation may occur before feedback is available (i.e. during planning or execution). This is a valuable avenue to explore for further research.

Estimating performance based on task-relevant cues

The second contribution to JoPs were from estimations based on task heuristics. The regression analysis suggested that cue-based estimation may be a secondary process, as binary predictors of changes in task parameters constituted a smaller contribution than task performance (PiZ scores). Constructing judgments from task-relevant cues has been a principle framework for metacognition (Koriat, 2000; Koriat, 2008), and weighted cue-based estimation has been proposed as a useful framework with which to understand driver self-assessments (Horrey et al., 2015). Our evidence of a cue-based metacognitive mechanism both validates Horrey et al.'s proposals, and the use of metacognitive principles for assessing awareness of perceptual-motor performance.

When participants received only visual feedback the cue-based estimation process appears to be driven by speed cues, rather than zone width cues. This agrees with recent findings that humans are sensitive to global flow speed when steering bends, even when flow speed information per se is not critical to the task at hand (e.g. maintaining position in lane; Kountouriotis, Mole, Merat & Wilkie, 2016), and leads to the intriguing possibility that the observed flow induced steering biases in Kountouriotis et al. (2016) could at least partly be due to metacognitive influences over steering control.

Whilst cue-based estimation may be an efficient heuristic, it relies on appropriate weightings that are learned over time. Initial cue weightings will be determined by prior experience, but the effective use of cue-based estimation needs prior cue-weights to be updated using feedback. The group receiving auditory feedback had different weightings attributed to each heuristic – suggesting feedback can be used to update cue weights - but this did not lead to increased monitoring resolution (in this particular task). The process determining how cue weightings (and cue-selection) are learned is an important open question that is of relevance to any situation where decisions are made based upon perceptual-motor performance.

Conclusions

The current manuscript presents an attempt to bring together cross-cutting themes in perception, action and cognition – combining perspectives from metacognition, vision science and traffic psychology to assess metacognitive awareness of perceptual-motor performance. Our results provide initial evidence that judgements are generally poor, but that the signals used to form judgements may vary depending on the task conditions. Suggestions of a dual-process account of judgment generation have been presented, consisting of: i) monitoring of perceptual-motor *performance* (Fleming et al., 2015; Sinanaj et al., 2015), and ii) *weighted cue-based estimation* (Koriat, 2000; Koriat, 2008; Horrey et al., 2015). Our results suggest that drivers may be making strategic decisions about trajectory planning based on fairly flawed estimates of actual steering performance. The question remains how to improve these estimates to help drivers gauge their own perceptual-motor capabilities. Whilst it seems likely that providing some form of additional feedback about errors could improve performance, our evidence does not suggest that this will necessarily allow drivers to better evaluate this performance. The task in the current manuscript is specific, and it is imperative that this model is tested in other perceptual-motor scenarios, under different monitoring constraints. Future research should also consider how these monitoring processes interact with non-perceptual priors not examined in the current manuscript – such as level of driver experience, or perceived competency.

References

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language*, *55*, 64–88. doi:10.1016/j.jml.2006.02.002

Amado, S., Arikan, E., Kaça, G., Koyuncu, M., & Turkan, B. N. (2014). How accurately do drivers evaluate their own driving behavior? An on-road observational study. *Accident Analysis and Prevention*, *63*, 65–73. http://doi.org/10.1016/j.aap.2013.10.022

Beilock, S. L., & Holt, L. E. (2007). Embodied Preference Judgements: Can Likeability Be Driven by the Motor System? *Psychological Science*, *18*(1), 51–57.

Billington, J., Field, D. T., Wilkie, R. M., & Wann, J. P. (2010). An fMRI study of parietal cortex involvement in the visual guidance of locomotion. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(6), 1495–507. http://doi.org/10.1037/a0018728

Carter, M. J., Smith, V., & Ste-Marie, D. M. (2016). Judgments of learning are significantly higher following feedback on relatively good versus relatively poor trials despite no actual learning differences. *Human Movement Science*, *45*, 63–70. http://doi.org/10.1016/j.humov.2015.11.006

Dunlosky, J. & Bjork, R.A. (2008). *Handbook of metamemory and memory*. East Sussex, UK: Psychology Press.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self- assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106. doi:10.1111/j.1529-1006.2004.00018.x

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–33. http://doi.org/10.1038/415429a

Field, D. T., Wilkie, R. M., & Wann, J. P. (2007). Neural systems in the visual control of steering. *The Journal of Neuroscience*, *27*(30), 8002–10. http://doi.org/10.1523/JNEUROSCI.2130-07.2007

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. http://doi.org/10.3758/PBR.15.1.174

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific Disruption of Perceptual Confidence. *Psychological Science*, *26*(1), 89–98. http://doi.org/10.1177/0956797614557697

Fourneret, P., & Jeannerod, M. (1998). Limited conscious monitoring of motor performance in normal subjects. *Neuropsychologia*, *36*(11), 1133–1140. http://doi.org/10.1016/S0028-3932(98)00006-2

Freund, B., Colgrove, L. A., Burke, B. L. & McLeod, R. (2005). Self-rated driving performance among elderly drivers referred for driving evaluation. *Accident Analysis and Prevention, 37,* 613-618.

Frissen, I. & Mars, F. (2014). The effect of visual degradation on anticipatory and compensatory steering control. *The Quarterly Journal of Experimental Psychology*, *67:3*, 499-507. DOI: 10.1080/17470218.2013.819518

Fuller, R. (2005). Towards a general theory of driver behaviour. *Accident Analysis and Prevention*, *37*(3), 461–472. http://doi.org/10.1016/j.aap.2004.11.003

Goszcynska, M. & Roslan, A. (1989). Self-evaluation of drivers' skill: A cross-cultural comparison. *Accident Analysis and Prevention, 21(3),* 217-224.

Gregersen, N. P. (2006). Young Drivers' Overestimation of their own skill – an experiment on the relation between training strategy and skill. *Accident Analysis and Prevention, 28(2),* 243-250.

Horrey, W. J., Lesch, M. F., Mitsopoulos-Rubens, E., & Lee, J. D. (2015). Calibration of skill and judgment in driving: Development of a conceptual framework and the implications for road safety. *Accident Analysis and Prevention*, *76*, 25–33. http://doi.org/10.1016/j.aap.2014.12.017

Jersakova, R., Souchay, C., & Allen, R. J. (2015). Negative Affect Does Not Impact Semantic Retrieval Failure Monitoring. *Canadian Journal of Experimental Psychology, 69*(4), 314–326. doi: 10.1037/cep0000065

Kountouriotis, G. K., Floyd, R. C., Gardner, P. H., Merat, N., & Wilkie, R. M. (2012). The role of gaze and road edge information during high-speed locomotion. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 687–702.

Kountouriotis, G. K., Mole, C. D., Merat, N. & Wilkie, R. M. (2016). The need for speed: global optic flow speed influences steering. *Royal Society: Open Science,* 3: 160096. http://dx.doi.org/10.1098/rsos.160096

Koriat, A. (2000). The feeling of knowing: some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*, 149-171.

Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory and Cognition, 36*(2), 416-428.

Koriat, A. & Bjork, R.A. (2005). Illusions of Competence in Monitoring One's Own Knowledge During Study. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(2), 187-194.

Kovácsová, N., de Winter., J. C. F., Schwab, A. L., Christoph, M., Twisk, D.A.M. & Hagenzieker, M. P. (2016). Riding performance on a conventional bicycle and a pedelec in low speed exercises: Objective and subjective evaluation of middle-aged and older persons. *Transportation Research Part F: Traffic Psychology and Behaviour. 42*. 28-43.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*, 742–744.

Land, M. F., & Horwood, J. (1995). Which parts of the road guide steering? *Nature, 377*, 339-340

Lappi, O. (2014). Future path and tangent point models in the visual control of locomotion in curve driving. Journal of Vision, 14, 1–22

Lorch, R. F., & Myers, J. L. (1990). Regression Analyses of Repeated Measures Data in Cognitive Research. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *16*(1), 149–157.

Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science*, *19*(6), 601–6. http://doi.org/10.1111/j.1467-9280.2008.02129.x

McRuer, D., & Allen, R. (1977). New results in driver steering control models. *Human Factors, 19 (4)*. 381-397

Metcalfe, J., Eich, T. S., & Miele, D. B. (2013). Metacognition of agency : proximal action and distal outcome. *Experimental Brain Research*, *229*, 485–496. doi:10.1007/s00221-012-3371-6

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & R. C. Schwing (Eds.), Human behavior and traffic safety (pp. 485–524). New York: Plenum Press.

Mole, C. D., Kountouriotis, G., Billington, J., & Wilkie, R. M. (2016). Optic Flow Speed Modulates Guidance Level Control: New Insights Into Two-Level Steering. *Journal of Experimental Psychology: Human Perception and Performance, 42*(11), 1818-1838. http://dx.doi.org/10.1037/xhp0000256

Nelson, T., & Narens, L. (1990). Metamemory: a theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–173.

Peters, M. A., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife*, 10.7554/eLife.09651. http://doi.org/http://dx.doi.org/10.7554/eLife.09651

Pisella, L., Gréa, H., Tilikete, C., Vighetto, a, Desmurget, M., Rode, G., … Rossetti, Y. (2000). An "automatic pilot" for the hand in human posterior parietal cortex: toward reinterpreting optic ataxia. *Nature Neuroscience*, *3*(7), 729–736. http://doi.org/10.1038/76694

Rhodes, M.G. & Castel, A.D. (2008). Memory predictions are influenced by perceptual information: evidence from metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615-25.

Roberts, S. C., Horrey, W. J., & Liang, Y. (2016). Measurement of driver calibration and the impact of feedback on drivers' estimates of performance. *Accident Analysis and Prevention*, *88*, 150–158. http://doi.org/10.1016/j.aap.2015.12.013

Robertshaw, K. D., & Wilkie, R. M. (2008). Does gaze influence steering around a bend? *Journal of Vision*, *8(4)*(18), 1–13. http://doi.org/10.1167/8.4.18.Introduction

Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. *Perception*, *33*(10), 1233–48.

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. http://doi.org/10.1016/j.cub.2011.10.030

Simons, D. J. (2013). Unskilled and optimistic: overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review*, *20*, 601–7. http://doi.org/10.3758/s13423-013-0379-2

Sidarus, N. & Haggard, P. (2016). Difficult action decisions reduce the sense of agency: A study using the Eriksen flanker task. *Acta Psychologica, 166*, 1-11.

Sinanaj, I., Cojan, Y., & Vuilleumier, P. (2015). Inter-individual variability in metacognitive ability for visuomotor performance and underlying brain structures. *Consciousness and Cognition*, *36*, 327–337. http://doi.org/10.1016/j.concog.2015.07.012

Sundström, A. (2008). Self-assessment of driving skill - A review from a measurement perspective. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(1), 1–9. http://doi.org/10.1016/j.trf.2007.05.002

Sundström, A. (2011). The validity of self-reported driver competence: Relations between measures of perceived driver competence and actual driving skill. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(2), 155–163. http://doi.org/10.1016/j.trf.2010.11.011

Wilkie, R. M., Kountouriotis, G. K., Merat, N., & Wann, J. P. (2010). Using vision to control locomotion: looking where you want to go. *Experimental Brain Research*, *204*(4), 539–47. http://doi.org/10.1007/s00221-010-2321-4

Wilkie, R. M., & Wann, J. P. (2002). Driving as night falls: the contribution of retinal flow and visual direction to the control of steering. *Current Biology*, *12*(23), 2014–7.

Wilkie, R. M., Wann, J. P., & Allison, R. S. (2008). Active gaze, visual look-ahead, and locomotor control. *Journal of Experimental Psychology. Human Perception and Performance*, *34*(5), 1150–64.