Article:
Engström, J, Bärgman, J, Nilsson, D et al. (4 more authors) (2018) Great expectations: A predictive processing account of automobile driving. Theoretical Issues in Ergonomics Science, 19 (2). pp. 156-194. ISSN 1463-922X

https://doi.org/10.1080/1463922X.2017.1306148

Great expectations: A predictive processing account of automobile driving

Johan Engström, Jonas Bärgman, Daniel Nilsson, Bobbie Seppelt, Gustav Markkula, Giulio Francesco Piccinini, Trent Victor

Predictive processing has been proposed as a unifying framework for understanding brain function, suggesting that cognition and behaviour can be fundamentally understood based on the single principle of prediction error minimization. According to predictive processing, the brain is a statistical organ that continuously attempts get a grip on states in the world by predicting how these states cause sensory input and minimizing the deviations between the predicted and actual input. While these ideas have had a strong influence in neuroscience and cognitive science, they have so far not been adopted in applied human factors research. The present paper represents a first attempt to do so, exploring how predictive processing concepts can be used to understand automobile driving. It is shown how a framework based on predictive processing may provide a novel perspective on a range of driving phenomena and offer a unifying framework for traditionally disparate human factors models.

Keywords: Predictive processing, expectancy, driving, driver behaviour, perception, action

Relevance to human factors / ergonomics theory: The objective of the paper is to explore how the predictive processing framework, which recently had a strong impact in basic neuroscience and cognitive science, can also offer novel perspectives on applied human factors problems, in this case related to automobile driving.

Introduction

In recent years, a novel framework for understanding brain function has emerged, suggesting that cognition and behaviour can be fundamentally understood in terms of the single principle of prediction error minimization. The key proposal is that that the brain is a statistical organ that attempts to "be in tune with", or "get a grip on", the world by continuously predicting its own sensory input and minimizing the deviations between the predicted and actual input. From a biological perspective, such a "grip" is needed to resist disorder (entropy), and maintain the body state (e.g., blood sugar level and temperature) within viable boundaries in the face of a constantly changing environment (Friston, 2010). While several variants of this general idea have been developed, Clark (2013, 2016) has suggested *predictive processing* as an umbrella term for these types of models which is also adopted in the present paper.

Traditional information processing accounts of cognition suggest an essentially feed-forward stream of processing (of information and/or neural activity) from sensory input, via perception and higher-level cognition, to action[1]. Predictive processing turns this view on its head, suggesting that the brain continuously generates predictions of its own sensory input and uses the actual online sensory signal mainly to check and correct the predictions (Clark, 2016). This re-framing promotes a view of cognition as expectation-driven active engagement with the world rather than as a set of sequenced information processing stages.

In recent years, these ideas have had a strong influence in both neuroscience and cognitive science in general (e.g., Bar, 2007; Clark, 2013, 2016; den Ouden, Kok, and de Lange, 2012; Friston, 2010; Friston et al., 2016; Hohwy, 2013; Huang, 2008) and models based on predictive processing concepts have been developed for a wide range of neural and cognitive phenomena such as perception, action generation, attention, learning, decision making, and emotion, often in the form of detailed mathematical formulations and computer simulations (e.g., Feldman and Friston, 2010; Friston, 2005; Friston et al., 2010; Friston et al., 2016; Rao and Ballard, 1999; Pezzulo, 2013; Pezzulo et al., 2015; Seth, 2013). More specifically, the framework has offered new perspectives on classical cognitive phenomena such as visual illusions (Brown et al., 2013), binocular rivalry (Hohwy, Roepstorff and Friston, 2008) and psychiatric disorders such as schizophrenia (Fletcher and Frith, 2009; Friston et al., 2014). A particularly attractive feature of the predictive processing framework is that it offers a unified account of sub-fields in cognitive science that have traditionally been studied more or less in isolation. As such, it has been proposed as a “unified theory of the brain” (Friston, 2010; Huang, 2008) or a “package deal” for cognitive science (Clark, 2013, 2016).

However, while existing human factors theories and models have often emphasized the key role of prediction and expectancy (e.g., Hollnagel, Nåbo and Lau, 2003; Endsley, 1995; Rasmussen, 1985; Summala, 1988; Vicente, and Rasmussen, 1992; Wickens et al., 2003), the more specific concepts offered by predictive processing have so far not been adopted in applied human factors research. The present paper represents a first attempt to do so, exploring the application of predictive processing in the context of automobile driving. The paper begins with a review of the predictive processing framework and its key underlying ideas. These concepts are then applied to a range of driving phenomena, including driver reactions to unexpected hazards, visual scanning at intersections and interaction with automatic interventions and automated driving functions. This is followed by a discussion of how the predictive processing framework relates to existing human factors theories and models. The paper ends with a brief discussion of some common themes emerging from the example applications and some suggestions for future development of these ideas.

Predictive processing

As already mentioned, predictive processing has been suggested by Clark (2013, 2016)

---

[1] It should be mentioned that several existing models in the information processing tradition do incorporate expectation and top-down processing as key features. This is particularly true for human factors models developed to account for operator behavior in real-world dynamic environments (e.g., Endsley, 1995, 2015; Rasmussen, 1985; see the review below). However, the basic concept of a feedforward chain of information processing through a set of distinct stages is still very prevalent both in cognitive science and applied human factors (see e.g., Wickens, 2002, Figure 2).

as an umbrella term encompassing several more specific accounts based on the common idea of the brain as a statistical prediction device. These include *predictive coding* (Rao and Ballard, 1999) and *active inference/the free energy principle* (Friston, 2010; Friston et al., 2010; Friston et al. 2016; see Huang, 2008, for a popular science account). A common source of these ideas is the work of von Helmholtz (1876), who viewed perception as a largely a top-down, constructive, process which can be understood in terms of probabilistic inference, akin to statistical hypothesis testing in science. More recently, this concept has been further developed in terms of the *Bayesian brain hypothesis* (e.g., Knill and Poguet, 2004). At the same time, as emphasised by Clark (2016), predictive processing also has close links to *embodied cognitive science*, emphasizing that cognition is crucially enabled and constrained by the physical body and its interaction with a structured environment. See Clark (2013, 2016) for more comprehensive reviews of the historical origins of current predictive processing models.

The central idea underlying predictive processing is that the key function of the brain is to capture statistical regularities in the environment and in the body by continuously trying to predict its own sensory input and acting on the resulting prediction errors (i.e., the mismatch between the predictions and the actual sensory input). Predictions can be thought of as generated by a *hierarchical generative model* embodied in the brain that has learned, over time, how states and events in the world, or one's own body, generate sensory input. The brain then attempts to minimise the deviation between predictions and actual sensory input in two principal ways:

- *Perception:* Updating the prediction based on the sensory input, and/or
- *Action:* Moving the eye, the head and/or the whole body to *bring about* the predicted sensory input (thus aligning it with the prediction).

In this way, the brain actively generates the expected sensory input (by perception and action) and, in case a prediction error occurs, updates the prediction or generates an action to cancel the error. Perception and action are thus closely intertwined, serving the common purpose of minimizing prediction error. A key notion that distinguishes predictive processing from most traditional accounts is thus that the brain processes and acts on *sensory prediction errors* rather than sensory input *per se*.[2] According to predictive processing, the brain is constantly engaged in such prediction error minimization, where predictions encompass the *complete* array of incoming sensory signals in all sensory modalities (what Clark, 2013, refers to as the "sensory barrage"). Thus, in every situation the brain generates predictions for the complete sensory input array, although in novel situations these predictions may be considerably off target (thus generating large prediction errors).

As these concepts may appear a little abstract at first, some concrete real-world examples may be useful before moving further. First, consider a driver following another vehicle on an otherwise empty highway. In this situation, the driver's generative model typically predicts that there should be no visual expansion (looming) of the lead

[2] This concept naturally explains some key features of brain anatomy and function that have been hard to fit within the traditional (feedforward information processing) view, including the abundance of backward ("top-down") connections between neural areas and the observed "default" activity in the brain observed at rest (Bar, 2007).

vehicle (a percept signifying "I am keeping a fixed distance gap"). If visual expansion is nevertheless registered by the visual sensory system (e.g., due to the lead vehicle slowing down), there will be a mismatch between predicted and actual looming, that is, a sensory prediction error. This looming prediction error can then be cancelled by updating the prediction (i.e., accepting that the lead vehicle is closing) and/or performing a braking action which cancels the looming, thus actively changing the looming input to align with the original prediction ("I am keeping a fixed distance gap").

It should be noted that the idea that the brain operates on prediction error rather than sensory input per se holds not only for lower-level sensory signals such as looming, but also for more abstract information at higher levels in the generative model. For example, consider a driver that uses (symbolic) speedometer readings to maintain the legal speed. Here, higher-levels in the generative model predict that speedometer readings will not exceed the legal speed limit. If this prediction is violated (i.e., the actual speed reading exceeds the speed limit), the prediction may be updated (perception) and/or an action may be generated to slow the vehicle down (to align with the high-level prediction that the legal speed will be maintained).

To explain the key principles underlying predictive processing in some further detail, we focus here on three principal mechanisms: (1) *Active inference* (accounting for the traditional concepts of perception and action), (2) *precision* (relating to the confidence in predictions but also to the traditional concept of attention) and (3) *model tuning* (relating to learning).[3]

Active inference

As mentioned above, a key notion in predictive processing is that the brain embodies a *hierarchical generative model* of how environmental (or bodily) states or events generate (or cause) sensory input (e.g., how a braking lead vehicle generates visual expansion on the retina, as in the example above). The generative model is hierarchical in the sense that it generates sensory predictions at different spatiotemporal scales, or levels of abstraction. Predictions at lower levels concern basic sensorimotor control (e.g., a predicted change in looming), whereas higher level predictions capture environmental (or bodily) regularities at larger scales (e.g., "we are approaching an intersection where the lead vehicle likely will slow down") and contextualise and disambiguate lower levels (see Pezzulo et al., 2015 for a good discussion). Another example of contextualisation is when a higher (context-level) prediction such as "I am about to drive through an intersection" helps to improve my lower-level predictions on

[3] Karl Friston and colleagues have developed a comprehensive mathematical framework for predictive processing known as the *free energy formulation* (e.g., Friston, 2010). Free energy is an information theoretic concept representing an upper bound on *surprise*, or *self-information* (formally defined as the negative log probability of an outcome). The general idea is that all cognition and behaviour could be understood based on the single mandate to minimise free energy (and thus surprise). Under certain simplifying assumptions, the minimization of free energy is equivalent to the minimization of prediction error, and, for present purposes, we will use this more intuitive conceptualization. Friston and colleagues have developed detailed simulation models based on the free energy principle which have been applied to explain a range of neuroscientific and psychological phenomena (see citations in the Introduction and in Clark, 2013, 2016).

where and when pedestrians are likely to enter the road. It is well-established that this hierarchical organization is reflected in the anatomy of the brain (e.g., Badre, 2008).

The term *active inference* refers to how agents infer environmental and bodily states from sensory input, and corresponds to the traditional notions of perception and action. Figure 1 provides a conceptual illustration of the key mechanisms underlying active inference in a hierarchical generative model. Here, the generative model is represented by a converging hierarchy of levels, which may be thought of as hierarchically organised brain regions[4]. At the bottom of the figure, environmental and body states cause three general types of sensory input signals: (a) *exteroceptive*, (b) *interoceptive,* and (c) *proprioceptive*. Exteroceptive inputs are generated by external (environmental) states such as the relative positions, velocities and accelerations of other road users and infrastructure elements. Interoceptive input is generated by visceral states, referring to the states of internal organs in the body (the viscera), what Damasio (1994) refers to as *emotions*. In driving, such states include the bodily representations of feelings of comfort or threat, for example related to the distance to a lead-vehicle in a car-following situation (see e.g., Summala, 2007). Finally, proprioceptive sensory input is generated by the states of muscles and joints. In driving, this would be, for example, the extension and contraction of muscles when depressing the brake pedal, the relative location of the shoulder joints when holding the steering wheel at a certain angle, or the rotation of the head and eyes. Of crucial importance, the brain does not have direct access to any of these states, but has to infer them based on sensory input.

As illustrated in Figure 1, the sensory input is continuously met by predictions at different levels in the generative model. At the lower levels such predictions concern detailed, modality-specific features such as visual looming and the shapes of objects while predictions at higher levels concern increasingly abstract and multimodal features extending over larger spatiotemporal scales such as “overtake”, “negotiate intersection” or, even more abstractly, “arriving on time”.

[4] Hierarchical architectures are very common in the cognitive- and brain modelling literature (e.g., Deco and Rolls, 2005; Cooper and Shallice, 2000; Engström, Victor and Markkula, 2013; Markkula, 2015). What the predictive processing account adds is a novel account of how the different levels interact with each other (in terms of predictions and precision-weighted prediction errors – see below), and how this process is governed by the single mandate to minimise prediction error (see e.g. Friston, 2005; 2008 and Rao and Ballard, 1999, for examples of more detailed hierarchical architectures for predictive processing).

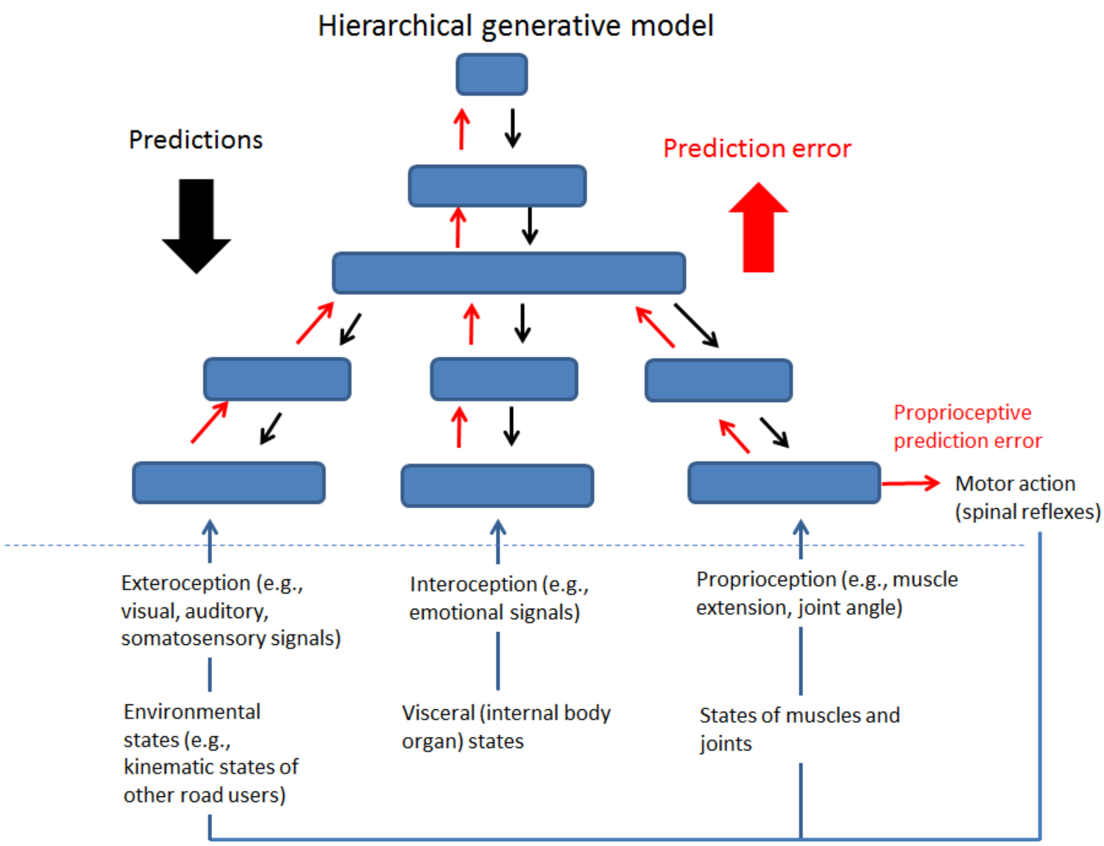

Figure 1: Conceptual illustration of inference in a hierarchical generative model, where sensory prediction errors are propagated upwards and met by predictions at each hierarchical level. By minimizing the global prediction error, the system settles into a new state, generates new predictions, and so forth. Actions are generated by cancelling proprioceptive prediction errors, that is, deviations between actual and predicted muscle extension and/or relative joint angle, as shown to the lower right (see text for further explanation).

When the input to a certain level in the hierarchy illustrated in Figure 1 does not match the prediction, a *prediction error* is generated. Active inference then amounts to cancelling this prediction error which, as noted above, is the fundamental joint role of perception and action: Perception here corresponds to updating the prediction to match the sensory input while action involves altering the sensory input to conform to the prediction, thus actively *bringing about* the predicted sensory input by means of movement (Clark, 2016). In most real-world situations, perception and action work closely in tandem with the common goal of minimizing prediction errors.

If the prediction error cannot be resolved at the present level of the generative model, it is passed upward for potential resolution at higher levels in the hierarchy. When a higher-level prediction is updated, it generates new descending (top-down) predictions which attempt to cancel prediction errors at the lower level. To illustrate this dynamic interaction between higher and lower levels in the generative model, consider a scenario where a driver approaches an intersection in a right-turn-only lane with a green light. Let’s further assume that our driver has failed to understand that he is in a right-turn-only lane and intends to go straight through the intersection. He thus enters the intersection on the assumption that he has the right of way. The driver then notices that other vehicles (unexpectedly to our driver) are honking at him.

Figure 2 illustrates a specific mechanism for how the resulting prediction error (unexpected honking) may be resolved through interaction between a lower (sensorimotor) level and a higher (context) level of the generative model, based on a greatly simplified version of the neural architecture proposed by Friston and colleagues (e.g., Friston 2005). The top panel represents the situation when the driver approaches the intersection with the context-level prediction that he will pass straight through the intersection in the correct lane. This generates various more specific top-down predictions of expected input at the sensorimotor level, in this case exemplified by "green light" and "no honking". At this point, these top-down predictions are still consistent with the actual sensory input (given that the driver failed to perceive any cues indicating that he is in the wrong lane), and the driver enters the intersection as planned. However, at this point other road users start to honk their horns to alert our driver on his mistake, generating a large prediction error (unexpected honking; middle panel) that is passed upwards. Based on this error, a new context-level prediction is generated that better fits the evidence, namely "I'm in the right turn lane going straight and that's why people are honking at me…". This represents the best hypothesis available to the driver that explains the honking input and thus reduces the overall prediction error back to zero (lower panel). In this way, only *unexplained* sensory input is passed forward/upward and acted upon, and perception involves a dynamic process of interaction between levels until reaching a stable state where the prediction errors at all levels in the hierarchy have been resolved in a satisfactory way. This yields a stable percept (and, at higher levels, a stable "context understanding" or situation awareness[5]) which, from a Bayesian perspective, represents the "best guess" of the external (or bodily) states-of-affairs given the available sensory data and the generative model. A detailed mathematical account of how this inference process might work is offered by Friston (e.g., 2005, 2010; Friston et al., 2016). This dynamic, interactive, process of prediction error minimization can also be seen as analogous to the more traditional notion of *evidence accumulation* as the basis for optimal decision making (e.g., Gold and Shadlen, 2001).

[5] See below for a discussion of the relation between predictive processing and the classical model of situation awareness (Endsley, 1995)

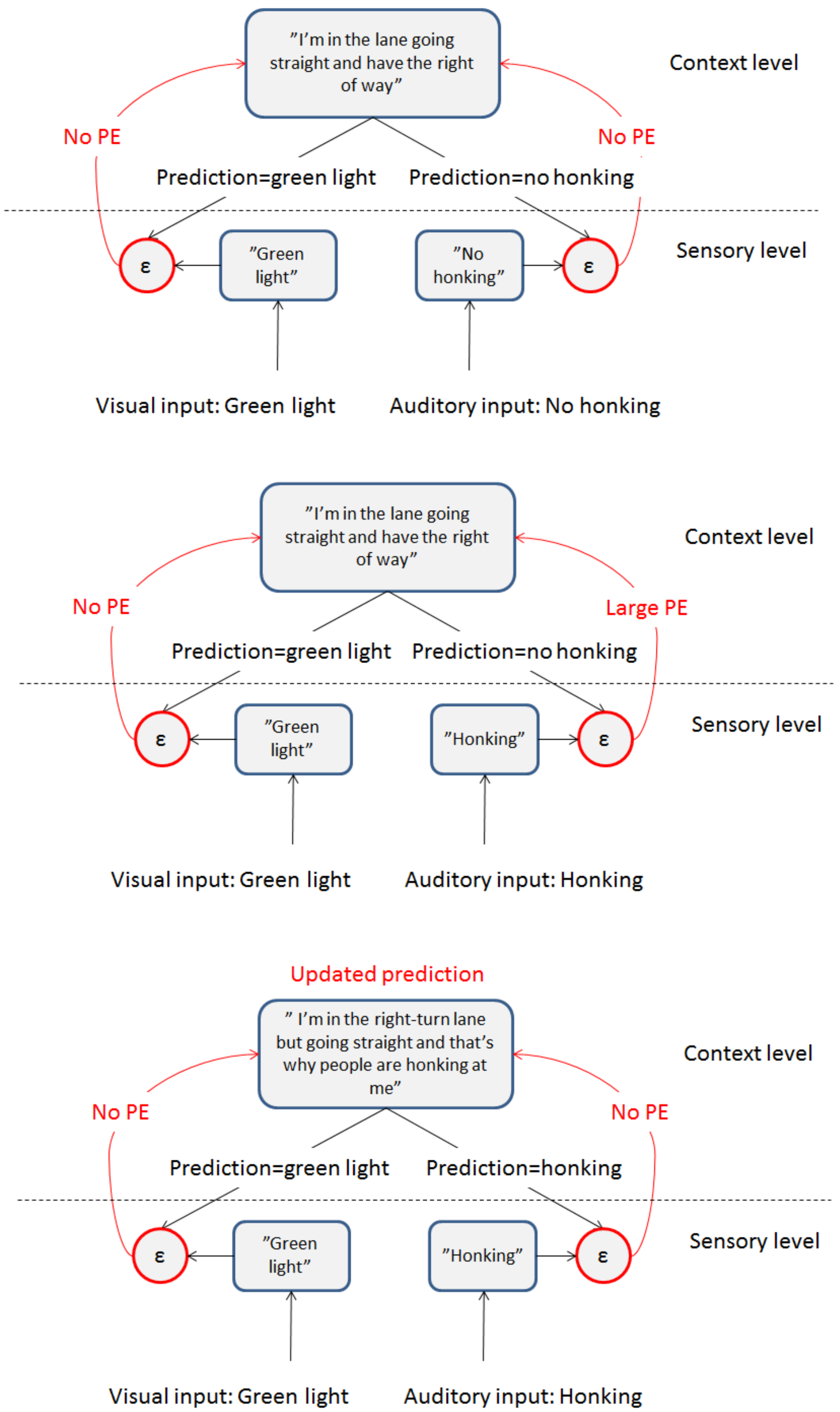

Figure 2: Illustration of a mechanism for how prediction errors may be resolved through interaction between levels in the generative model. This represents a greatly simplified version of the architecture proposed by Friston (e.g., 2005) with separate "state" units representing predictions (the grey boxes) and prediction error units (ε) representing the deviation between higher-level and lower level predictions, passing the prediction error to the higher level. See the text for a description of the example scenario.

Predictive processing also offers a new perspective on how actions are *generated*. Conventional views on motor control suggest that actions are generated by the brain sending action commands to the muscles via the spinal cord (e.g., Wolpert and Miall, 1996). By contrast, the predictive processing account suggests that the generation of action at the muscular level can be understood in terms of the cancelling of proprioceptive prediction errors. As illustrated at the bottom-right in Figure 1, proprioceptive prediction errors are generated through top-down predictions from higher levels such as “I will now press the brake”. If I am not yet braking (and the muscles and joints remain in their original position), the prediction will thus generate a proprioceptive prediction error at the spinal level that will automatically be cancelled by peripheral reflexes (such as the stretch reflex), thus generating the braking action. Viewed from this perspective, actions thus become “self-fulfilling prophecies” where predictions become equivalent to intentions (see Friston et al., 2010, and Adams, Shipp and Friston, 2013, for detail).

Precision

A key further notion in the predictive processing framework is that predictions from the generative model are *probabilistic*, thus representing the probability of sensory data given their causes. Thus, predictions (at every level in the hierarchical generative model) also involve an estimation of the expected precision (the inverse variance, or certainty) of the sensory input. This precision estimate then scales the resulting sensory prediction errors such that deviations resulting from certain (high-precision) predictions are given a higher weighting than prediction errors related to low-precision predictions. Thus, the ascending (bottom-up, upwards-flowing) prediction errors that drive active inference (perception and action) for a certain task are *precision-weighted* so that trusted sensory input is given a relatively stronger influence. This is analogous to statistical inference in hypothesis testing where, for example, the t-statistic (representing how much an experimental result can be trusted) is obtained by dividing the observed group differences with the estimated population variance. Thus, the t-statistic can be viewed as analogous to a precision-weighted predicted error (Feldman and Friston, 2010).

For example, when driving in reduced visibility conditions (e.g., fog or rain) it makes sense to selectively base corrective steering actions on those aspects of the visual scene that are reliably perceived (e.g., lane markings close to the vehicle) rather than on those harder to perceive in the adverse conditions (e.g., the curvature of the roadway in the distance). In this way, the precision-weighting of prediction error can be viewed as a means for selective enhancement of sensory input, and thus related to the traditional notion of attention (Feldman and Friston, 2010). Feldman and Friston (2010) further suggest that precision weighting may be implemented in the brain by means of established mechanisms such as neural synchrony and/or classical neuromodulation (e.g., transient changes in synaptic efficacy through the effects of e.g., dopamine or acetylcholine).

The concepts of precision and precision weighted prediction error are illustrated in Figure 3. Here, environmental or bodily states and events generate sensory input with a certain probability distribution, illustrated at the bottom of the Figure. The generative model predicts these sensory inputs with a level of precision, as represented by the probability distribution at the top. To match the sensory input, the model predictions thus need to account for both the mean and the variance (over time) of the sensory input. For each sensory input and prediction, the resulting prediction error is weighted by the precision associated with the prediction.

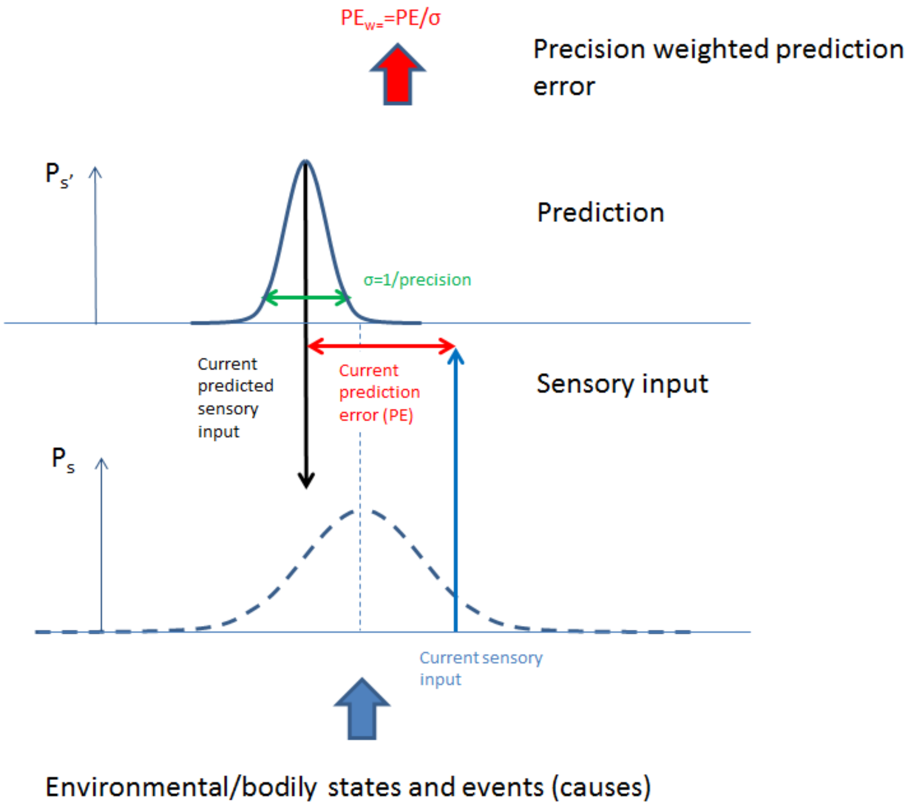

Figure 1 Illustration of the concepts of precision and precision weighted prediction error.

As discussed at length by Clark (2016), precision weighting of prediction error also provides a more general mechanism for controlling behaviour, or "sculpting the flow" in the generative model beyond the selective enhancement of sensory input discussed so far. In particular, this mechanism may be used to shift the balance between higher and lower levels in the generative model in the control of behaviour. Shifting control to higher levels means increasing the trust in your knowledge relative to the actual sensory input. Thus, by increasing the precision of prediction errors at higher levels (relative to lower levels) in the generative model, behaviour may be controlled in a more flexible manner, driven top-down by longer-term predictions, relating to the traditional notion of *cognitive control* (Miller and Cohen, 2001)[6]. This may be applicable in situations with high levels of uncertainty (e.g., driving through a complex intersection), when overriding lower-level, automatized behaviours in order to adapt to novel contexts (e.g., looking right instead of left when crossing the street in the UK) or achieve more distal

[6] This idea is in line with the extensive literature suggesting a key role of dopamine in "gating" access to higher level brain areas such as the pre-frontal cortex (e.g., Braver and Cohen, 2000).

goals (e.g., reduce traveling speed to improve fuel efficiency). The same mechanism may also be used for blocking actions during planning; by reducing the weighting of prediction errors at the proprioceptive level, actions may be contemplated without actually being executed (Clark, 2016).

Finally, as discussed in Clark (2016) and Friston et al. (2015), precision plays a key role in the arbitration between *pragmatic* (or extrinsic) and *epistemic* (knowledge-oriented) action, relating to the classical distinction between *exploitation* and *exploration* (e.g., Aston, Jones and Cohen, 2005). For example, consider that you are trying to find your way out in a dark room. Walking towards the exit is the pragmatic action that fulfils the main current prediction (that you will leave the room). If the room is familiar, you may be able to walk directly to the exit based on memory. However, in case you are uncertain of your location or the layout of the room, you may need to sense your way by touching the walls with your hands, which is an example of epistemic action. Thus, epistemic actions (e.g., in a driving context, checking for oncoming cars before overtaking) serve to increase the precision of the main current prediction (e.g., "I will overtake successfully"). In circumstances where the agent is confident in her predictions (i.e., the predictions are associated with high precision), prediction errors are usually most effectively cancelled by carrying out the predicted action directly (i.e., pragmatic action). For example, consider a scenario where a driver looks back to the road to find that the vehicle is slightly off the predicted path. This driver may assess the situation with a high-degree of confidence (precision) and immediately cancel the prediction error with a steering correction. However, if the prediction is associated with high uncertainty (low precision), it is generally better (from a global prediction error minimization standpoint) to first perform epistemic actions (e.g., visual sampling) to increase precision before performing the predicted pragmatic action. For example, if our driver looks up and discovers that a lead vehicle is slowing down (on a multi-lane motorway) she may generate a new prediction that entails overtaking the lead vehicle. However, this prediction is associated with some uncertainty since other vehicles may approach from behind. Thus, a mirror check (an epistemic action) is needed to increase precision before initiating the overtaking (it may be noted that, in this example, the concept of "prediction" becomes almost synonymous with "decision").

Model tuning

This third principal mechanism underlying the predictive processing framework relates to the modification of the generative model itself, with the general purpose to minimise prediction error in the long term, thereby *tuning* the brain to behaviourally relevant statistical regularities in the environment. Thus, in contrast to (perceptual) inference which, as described above, deals with the moment-to-moment update of predictions and their precision based on incoming sensory data, model tuning occurs more gradually over longer time scales. When this occurs during the organism's lifetime, it corresponds to the traditional concept of learning[7].

---

[7] Model tuning also occurs over evolutionary time scales (through natural selection) but, since this is less relevant for driving, it is not further addressed here.

Learning, according to predictive processing, involves a kind of "bootstrapping" (see Clark, 2016) where the novice's crude attempts to predict the sensory input initially generates large prediction errors. By gradually adjusting model parameters (e.g., synaptic strength at the neural level; Hebb, 1949) to improve predictions, the model will begin to capture behaviourally relevant regularities in the sensory input. If these regularities are consistently and frequently encountered, a strong model will eventually develop yielding skilled, fluent behaviour. This has a close correspondence with traditional accounts of *reinforcement learning* (e.g., Rescorla and Wagner, 1972; Sutton and Barto, 1998; Montague, Hyman and Cohen, 2004), where model parameters are gradually adjusted in order to maximise expected reward or utility. For example, in temporal difference (TD) learning, encountering unexpected reward (as indicated by reward prediction error) leads to reinforcement of the behaviour that produced the reward by strengthening the model parameters associated with that behaviour. Conversely, if a behaviour does not lead to the expected reward, the corresponding model parameters are weakened. However, the concepts of reward, value or utility can be (conceptually and mathematically) subsumed under the more general framework of prediction error (or free energy) minimization (Friston, et al., 2015). From this perspective, valuable states are simply those states expected to minimise prediction error. For present purposes, the key point is that learning, according to predictive processing, involves a gradual modification of the generative model which improves predictions (across time-scales) and thus *tunes* the agent to behaviourally relevant statistical regularities in the world reflected in the sensory input (see e.g., Friston, et al., 2015 and Pezzulo et al., 2015, for more detailed discussions of these mechanisms).

Those statistical regularities in the sensory input that occur consistently and frequently will eventually result in strong, high-precision predictions with highly-weighted prediction errors, likely to be directly perceived and/or acted upon. In this situation, one could say that the agent has become *attuned* to the relevant statistical regularities in the world. Furthermore, according to the free energy formulation, what is minimised during learning is actually prediction error *plus* model complexity, as discussed in detail by Friston (2010) and FitzGerald et al., (2014) based on concepts from model selection in mathematical statistics. Thus, as suggested by Clark (2016), successful learning will result in minimally complex (cheap and frugal) models which are nevertheless sufficient to "do the job". This speaks to the classical notion of bounded rationality (Simon, 1955), or *satisficing*, which has become a key concept also in driver behaviour research (Boer, 1999; Goodrich, Stirling and Boer, 2000; Summala, 2007). A good example of this in the driving context is the task of lane keeping (in benign conditions), which is generally governed by highly consistent and frequent sensory input (e.g., lane markings and optical flow patterns) with predictable sensory consequences of corrective actions (e.g., steering corrections). Lane keeping is thus quickly learned and governed by a simple but strong generative model, yielding high-precision predictions. This notion is supported by the fact that lane keeping is typically not negatively influenced by secondary tasks imposing a high cognitive (or working memory) load (Engström, 2011; Engström et al., in press).

The predictive processing framework thus offers a straightforward account of *automaticity*, which traditionally refers to behaviours that are consistently triggered by specific stimuli and for which task performance is unaffected when simultaneously performing another, secondary, task (Shiffrin and Schnedier, 1977). As suggested by Shiffrin and Schnedier (1977), automaticity develops through repeated exposure to consistent sensory-motor mappings in the task environment. As sketched above, the predictive processing framework offers a similar but more detailed account for how automaticity develops in terms of the gradual tuning of a generative model to statistical regularities in the sensory input, driven by the criteria of minimizing prediction error and model complexity. If the generative model tuning is viewed in terms of altering synaptic strength (Hebb, 1949), this view becomes closely aligned with existing accounts where automaticity is modelled in terms of neural pathway strength (e.g., Cohen et al., 1990; an idea applied in the driving context by Engström et al., in press).

Summary

According to the predictive processing framework, the brain is viewed as a statistical organ continuously trying to "get a grip" on, or "get in tune" with, statistical regularities in the world (including its own body) through sensorimotor interaction. The brain does this by constantly predicting sensory input and minimizing the resulting sensory prediction error (both instantly and over time).

The predictive processing framework, as outlined here, involves three principal mechanisms governing this on-going minimization of sensory prediction error: First, *active inference* refers to the moment-to-moment minimization of prediction error, which may be achieved by updating the prediction (perception) and/or changing the sensory input to conform to the prediction through motor activity (action). Perception and action are thus two sides of the same coin serving the joint purpose of minimizing prediction error. Second, prediction errors are weighted by the *precision* (certainty) of the corresponding prediction. This mechanism is related to the traditional notion of attention, and can be used to control behaviour in several different ways, including the selective enhancement of sensory input, altering between bottom-up and top-down driven behavioural control by shifting the relative weighing of prediction error between hierarchical levels in the model, and managing the trade-off between pragmatic and epistemic actions. Pragmatic actions refer to actions that directly cancel prediction error while epistemic (or knowledge-seeking) actions, such as explorative eye movements, serve the purpose to increase precision of predictions related to pragmatic action, thus minimizing the overall prediction error over time. Finally, the parameters of the generative model are *tuned* over time. By constantly predicting the sensory input, and gradually adjusting model parameters to yield better predictions (with less prediction error), the model eventually learns to capture statistical structures in the sensory input and, on this basis, infer external (and bodily) causes of the sensory input. Frequent exposure to consistent environmental regularities leads to less complex models yielding high-precision predictions. This suggests a mechanism for the development of automaticity.

Application to driving

In this section, we explore how the predictive processing framework outlined above may be applied in the context of automobile driving. We begin by considering how the framework may offer a more precise characterization of classical hierarchical models of driving (e.g., Michon, 1985). We then discuss a number of more specific applications, starting with drivers' emergency responses to a braking lead vehicle, focusing in particular on how observed differences in responses (and the lack of response) may be explained in terms of the extent to which the braking event was initially predicted by the driver. We then address an intersection scenario and consider how the present framework may account for visual scanning patterns and different types of expectation "failures". Finally, we discuss how predictive processing concepts may help to understand drivers' interaction with automatic steering interventions and automated driving functions. These examples are intended to provide a first illustration of how the proposed framework can improve our understanding, and generate testable hypotheses of different aspects of driver behaviour, and encourage its application to other driving-related phenomena.

Characterizing hierarchical levels of driving

One of the most persistent ideas in driving research is to characterise the driving task in terms of a hierarchy of goals or sub-tasks (Allen, Lunenfeld and Alexander, 1971; Michon, 1985). We will here adopt the terminology of Michon (1985) who distinguished between (1) the *operational* level, related to the momentary control of the vehicle (e.g., steering and braking), (2) the *tactical* level, relating to the current driving situation (e.g., selecting speed or deciding to overtake) and (3) the *strategic* level, relating to general goals of driving, such as destination and route choice.

The proposed framework potentially offers a more precise characterization of these levels which will be useful when addressing the specific applications below. In predictive processing terms, driving can generally be understood as *active inference* related to a set of driving-related sub-tasks. As outlined above, active inference entails continuously generating predictions and cancelling prediction errors by means of perception and action. Such inference can occur at different hierarchical levels of the driving task.

At the operational level, predictions concern sensory input relating to, for example, the momentary evolution of speed, headway, heading as well as proprioceptive input related to motor actions (e.g., steering, accelerator and brake pedal inputs). Operational driver behaviour can thus be understood as the minimization of prediction errors relative to those predictions, for example, by means of perceiving lane keeping deviations (by updating the prediction based on the prediction errors) and/or cancelling the error directly by steering corrections.

At the tactical level, predictions and prediction errors concern how the current situation will develop, for example, "I will change lane and then exit the road". Top-down predictions from the tactical level contextualises the operational level. The resulting prediction errors then contribute to the updating of predictions at both the operational and tactical levels (thus, in a sense, being *on* the operational control loop; see Gibson et al., 2016). A simple example of such interaction between higher and lower levels was illustrated in Figure 2 above.

Finally, predictions at the strategic level (being *on* the tactical control loop) concerns driving-related states on even longer spatiotemporal scales, such as the end destination and how to get there. This contextualises the tactical level by generating top-down predictions about expected traffic situations (e.g., "I will travel during rush hour and thus traffic cues ahead are likely to occur"). Of course, a more fine-grained set of levels may also be considered which, for example, distinguishes between sub-levels of operational control (e.g., between actions driven purely by spinal reflexes and those related to higher-level sensorimotor control).

Responding to a braking lead vehicle

In this section, we address the operational-level task of responding to a braking lead vehicle in a critical rear-end scenario, which, as we shall see, may also be strongly influenced by higher-level predictions. Traditionally, drivers' emergency avoidance reactions in critical situations have been mainly considered based on the concept of *reaction time*, or more precisely, *perception-response time* (Green, 2000; Olson, 1989; Olson and Sivak, 1986). Reaction time is usually considered a property of the individual driver influenced by various factors such as, for example, age, expectancy, and cognitive load (Green, 2000). However, recent analyses, both on driving simulator data (Ljung Aust, Engström and Viström; 2011) and naturalistic crash/near-crash data (Victor et al., 2015; Markkula et al, 2016) have shown that the timing of driver reactions in unexpected emergency situations are mainly determined by the situation kinematics rather than driver-specific perception-response times. In rear-end scenarios, the situation kinematics are largely reflected by the optical expansion (looming) of the lead vehicle. For example, the quantity τ (tau, the optical angle subtended by a front object, θ, divided by the angular rate, $\dot{\theta}$) provides an optically specified estimation of time-to-collision, and has been proposed as a key source of information that guides braking (Lee, 1976). Markkula et al. (2016) showed that drivers involved in rear-end crashes or near crashes typically did not initiate braking, or exhibit a physical response, until looming (here quantified as $1/\tau$) reached a certain level.

These accounts are all based on the traditional assumption that what drives the braking response is the sensory signal (e.g., looming) itself. Predictive processing adds a new twist by suggesting that drivers rather act on sensory (in this case looming) *prediction error*. In other words, this suggests that drivers in emergency rear-end situations react to *unexpected* (unexplained) looming rather than to looming *per se*[8]. As we shall see, this brings a new perspective that emphasises aspects of the braking response process that have been difficult to conceptualise based on the traditional view.

A predictive processing account of avoidance responses to looming suggests that the generative model constantly produces predictions of expected looming (or the lack of looming) and that drivers initiate avoidance actions (by braking and/or steering) when the looming prediction does not match the actual looming, with the purpose to eliminate the discrepancy. Importantly, such predictions may originate both from the operational and tactical levels of the generative model. First, consider a scenario where a driver is following a lead vehicle and the lead vehicle suddenly brakes in a manner unexpected to the driver. The driver's generative model thus initially predicts zero looming. In this special case, the looming prediction error simply reflects the looming input and the

[8] The same argument can be made for the brake light signal, although the brake light seems to play a less significant role than looming in determining driver reactions in real rear-end crashes and near crashes (Markkula et al., 2016).

predictive processing perspective does not really bring anything new to the table with respect to explaining when the driver will initiate braking[9]. Now consider the alternative scenario pictured in Figure 2 (based on a real near crash scenario in the naturalistic crash/near crash dataset used in Engström et al., 2013, and Markkula et al., 2016). Here, the driver approaches a truck about to exit the road. Despite strong looming indicating that a collision may be imminent, the driver does not slow down, presumably based on the assumption that the truck will be off the road well in time before she reaches the intersection. In terms of the present framework, the generative model predicts that looming will occur initially but that it will eventually go away (when the truck exits). This prediction is the result of exposure to many similar situations in which the lead vehicle did exit the road as expected, which has tuned the generative model to these types of scenarios. Thus, since the predicted looming (at the operational level) initially matches the actual looming, there is initially no prediction error that needs to be cancelled and no avoidance action ensues, despite strong looming cues being present. However, on this particular occasion, the exiting vehicle is blocked by another vehicle (not detected by our driver) and thus stopped in the middle of the road. Thus, at a certain point the actual looming is no longer predicted and will hence generate a strong prediction error that the driver will react upon (by steering and/or braking) in order to avoid a crash.

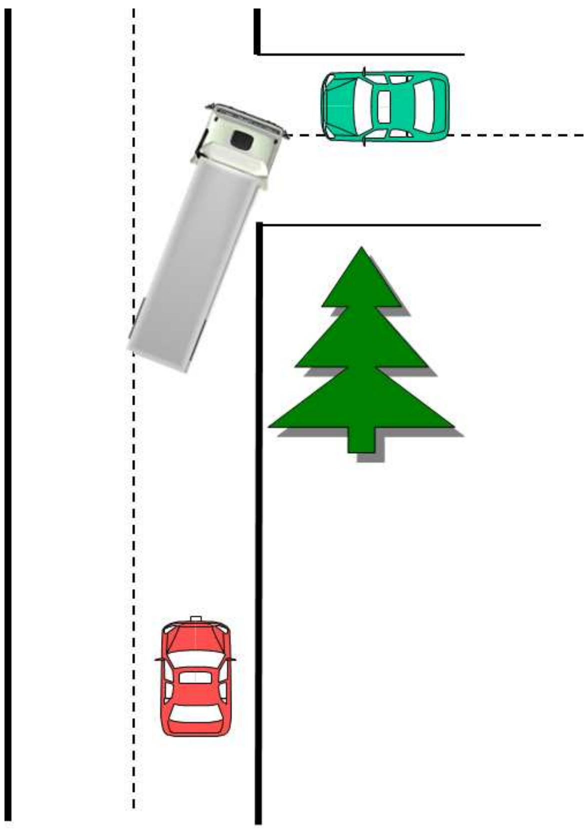

Figure 4: Example scenario where looming is initially expected.

[9] However, as demonstrated by Markkula (2014), a driver model acting on looming prediction error may be useful for reproducing on-going braking control.

This example shows how mismatching expectations may lead to critical situations when the driver erroneously predicts (based on previous exposure and the resulting tuning of the generative model) that the scenario will develop in a certain way and initially acts based on this prediction. A similar case is represented by a scenario, observed in three out of thirty heavy vehicle crashes in analysed in Engström et al. (2013; a more detailed analysis can be found in Engström, Bärgman and Lodin, 2016), where a driver enters an on-ramp to a motorway with a stopped vehicle ahead. These drivers kept up their speed to be able to merge onto the motorway, and initially accepted strong looming, presumably based on the expectation that they would eventually be able to merge. When dense traffic on the motorway prevented merging, the crash was unavoidable.

Based on traditional concepts (such as perception-response time) it would be easy to overlook this interpretation and misinterpret the type of scenario exemplified in Figure 2 (or the on-ramp crashes just discussed) as being due to a delayed driver response, caused, for example, by cognitive load or "looked-but-failed-to-see" phenomena (e.g., Koustanaï et al., 2008). The present account, suggesting that drivers react specifically to looming prediction error, offers a natural way to conceptualise this type of phenomenon and distinguish it from truly delayed reactions, related, for example, to driver distraction. It would be interesting to conduct a detailed analysis of a larger set of naturalistic crashes and near-crashes to further investigate the actual prevalence of the crash causation mechanism implied by the examples above.

Negotiating an intersection

Now, consider the scenario illustrated in Figure 5 where the driver intends to turn right at a non-signalised intersection, with a crossing bicycle lane. The tactical-level prediction in this case can be understood as something like "I will perform a successful right turn at a T-junction".

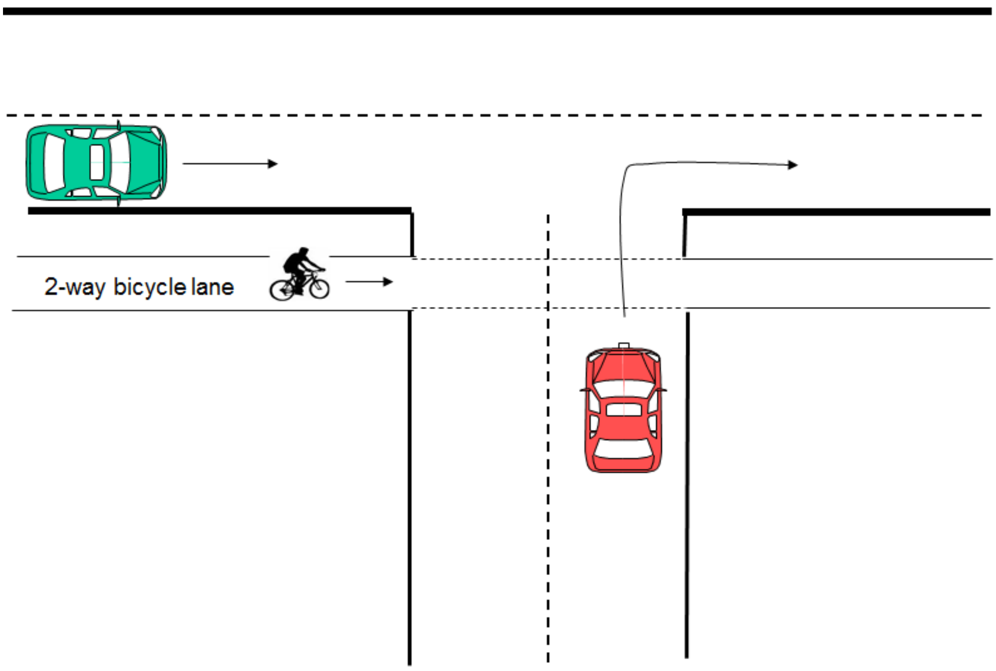

Figure 5: Example intersection scenario: Right turn at non-signalised intersection.

While turning, the generative model will continuously produce operational-level predictions of *all* sensory inputs expected to occur during the turn. This includes exteroceptive signals, in particular visual signals that guide heading. For example, as described by Wann and Land (2000), steering along a curved path towards an intended target could be achieved by fixating the target and then keep velocity and the rate of change of the visual angle to the target constant. Sensory predictions also concern proprioceptive signals (e.g., the relative position of joint angles in the arms and legs) and interoceptive signals (e.g., emotional signals from the body representing comfort/discomfort). Some hypothetical examples of predicted and actual sensory input time series at the operational level are illustrated in Figure 6. The figure also illustrates how the precision of the visual guidance prediction (here the predicted visual angle rate relative to the target location) is continuously updated based on incoming sensory data. In this example, the precision of the visual guidance prediction is reduced when looking away from the forward roadway (e.g., to scan for hazards to the left, as indicated by the glance signal) and increased again when looking back.

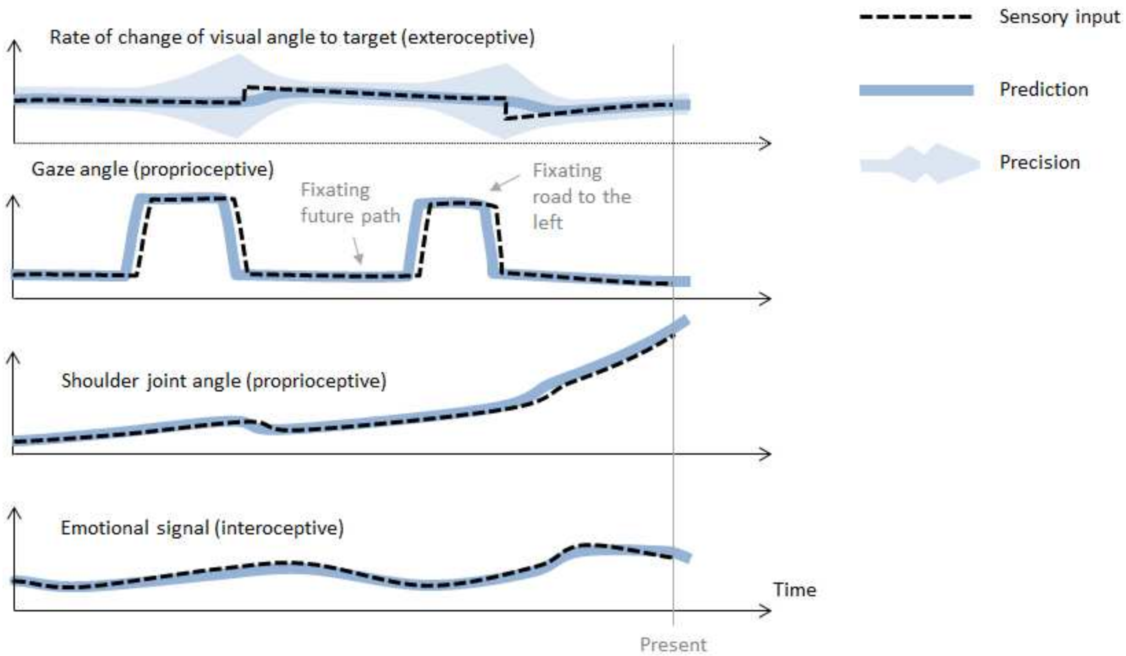

Figure 6: Conceptual illustration of sensory predictions as a multidimensional time series continuously updated based on and actual sensory input. Precision is only illustrated in the top graph, showing how the confidence bounds around the prediction of steering-related visual input are reduced as the result of visual scanning away from the future path (second graph from the top). Note that the predictions have been shown as extending just immediately beyond the present time, to emphasize that the main point of the predictions is not to say something about a distant future, but rather to continuously provide an estimate of what the next sensory input will be (thus, in a sense, “predicting the present”; Clark 2016).

The driver then *acts to fulfil the predictions*. At the operational level, this entails generating vehicle movements to sample sensory input that matches the prediction (Friston et al., 2010). For example, in terms of the steering model cited above (Wann and Land, 2000), this involves predicting constant velocity and angular rate and then controlling the vehicle in a way that generates sensory input fulfilling this prediction. As described above, actions, according to the proposed framework, are generated by cancelling proprioceptive prediction errors by peripheral reflexes. Thus, since the tactical-level prediction in this case is a successful right turn, the operational-level proprioceptive predictions entail changes in arm joint angle and muscle contraction that would result from turning the steering wheel to the right while keeping velocity and angular rate constant. When these proprioceptive predictions do not match the actual (sensed) proprioceptive signals (obtained from proprioceptors in the joints and muscles), the prediction error is smoothly cancelled by spinal reflexes contracting the relevant muscles to produce the steering action. Specifically, note in Figure 6 that, when looking back to the future path, the driver detects a small deviation from the desired constant visual angle rate, and responds to these by modifying the on-going prediction of shoulder joint angle.

Thus, successful turning at the intersection involves a tactical-level prediction ("I will perform a successful right turn at a T-junction") generating operational-level exteroceptive, proprioceptive, and interoceptive predictions illustrated in Figure 6 (see also Figure 1). The tactical-level prediction (which is not included in Figure 6) is initially associated with a great deal of uncertainty since both cars and bicyclists are relatively likely to cross the driver's path. Thus, epistemic action (visual sampling) is needed to reduce uncertainty (increase precision) before the turn can be performed. As suggested by Friston et al. (2012), predictions may be conceptually viewed as *hypotheses* while epistemic actions (e.g., visual sampling) can be thought of as *experiments* testing these hypotheses. It thus makes intuitive sense to sample those locations likely to provide information of relevance to the prediction (the hypothesis) at hand, just like it makes sense for the scientist to conduct experiments that confirm or disconfirm the general hypothesis that motivated the research. In any given situation, there is an optimal trade-off between pragmatic and epistemic action (i.e., between exploitation and exploration) and, as shown by Friston et al. (2015, 2016), this optimal trade-off is defined by the single mandate to minimise overall (across abstraction levels and over time) sensory prediction error (or, in their terms, minimize free energy). Hence, the tactical right turn prediction will include a mix of pragmatic and epistemic actions at the operational level which, for the experienced driver, represents a (near-) optimal prediction error minimization in this scenario. Thus, in a sense, a driver attuned to this intersection scenario expects that he will turn right *and* scan the road for hazards in a certain way, generating a time series of exteroceptive, proprioceptive and interoceptive sensory predictions which are continuously compared to the actual input (as illustrated in Figure 6). The observed driver behaviour can then, according to predictive processing, be fundamentally understood in terms of attempts to minimise the overall prediction error by means of perception and action.

If the scenario consistently and frequently plays out in a certain way (e.g., other road users almost always appear from the left), the gradual tuning of the generative model will eventually result in a simple but strong model generating more or less automatized scanning patterns and vehicle control actions. This has two important implications: First, for a driver that is well attuned to the scenario, visual scanning patterns would be expected to approximate an optimal trade-off between pragmatic and epistemic actions (based on the common mandate to minimise overall sensory prediction error). Hence, these patterns largely depend on (and can, at least in theory, be predicted from) the statistical properties of the traffic/roadway environment. While the account sketched here is purely conceptual, the mathematical framework outlined by Friston et al. (2015, 2016) may possibly be used to predict drivers' scanning patterns based on known scenario statistics. As further discussed below, this has interesting parallels with existing information-theoretic models of visual sampling (Senders, 1964, 1983; Kujala et al., 2016) and the SEEV (Salience, Effort, Expectancy, Value) model (Wickens et al., 2003).

A second implication is that problems may occur if the statistical regularities of the environment (that drivers' generative models are attuned to) suddenly change. A classic example of this type of situation is demonstrated by the work of Summala and colleagues (Summala and Räsänen, 2000; Räsänen and Summala, 1998), who set out to identify underlying causal mechanisms behind car-bicycle crashes at intersections and roundabouts in Sweden and Finland. In the 1990's, two-way bicycle lanes were becoming increasingly common in these countries and crash records indicated that car-bicycle crashes were more frequent at right turns with the bicycle approaching from the right than in other scenario configurations (Räsänen and Summala, 1998). Based on roadside observation of drivers' visual scanning behaviour, the authors suggested that these car-bicycle crashes often occurred due to the erroneous expectation that hazards would only approach from the left, which led to insufficient scanning to the right. As a result, bicyclists coming from the right, who had right of way and believed that the driver saw them, appeared outside the drivers' field of view and thus failed to capture the drivers' attention (Summala and Räsänen, 2000).

The present framework offers a straightforward interpretation of the origin of such expectancy problems: The Swedish and Finnish drivers' generative models had, over the years, been shaped by repeated exposure to right-turn situations where other road users typically appeared from the left. In this situation, looking to the right does not reduce uncertainty or overall prediction error much. Hence, with experience, drivers developed strong generative models yielding high-precision predictions that strongly biased automatized visual scanning towards the left. However, when two-way bicycle lanes were introduced, and bicyclists more frequently started to appear from the right, the drivers' generative models were suddenly "out of tune" with the statistical regularities of the environment, inducing expectancy mismatches when bicyclists appeared from the right which, in some cases, led to crashes. This represents a particularly problematic situation in which the drivers' predictions are inaccurate but still associated with high precision (i.e., you are in fact wrong but strongly believe that you are right).

This suggests a way to conceptualise expectancy mismatches in driving in terms of how the generative model is tuned to statistical regularities in the driving environment. Below, some prototypical cases are briefly discussed and illustrated by the type of representation introduced in Figure 3.

Inaccurate model and probable event

This corresponds to the mechanism observed by Summala and Räsänen (2000) just discussed, and is illustrated in Figure 7. Here, the generative model yields high-precision predictions in the current scenario, but the model is inappropriately tuned to the actual environmental statistics of that scenario. Thus, a crash may be generated by a relatively common event such as a bicyclist coming from the right, which the poorly tuned generative model fails to predict.

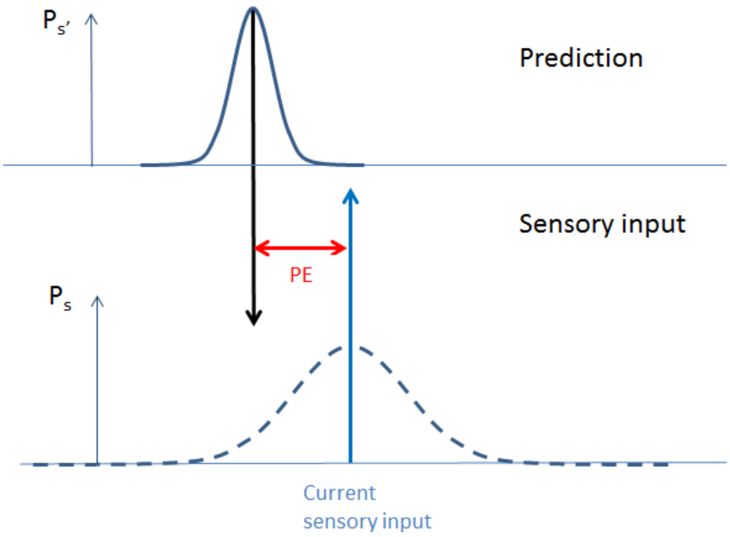

Figure 7: Inaccurate model and probable event.

Accurate model but improbable event ("bad luck")

This represents a situation where the generative model is well-tuned to the environment but the event is highly unlikely, as illustrated in Figure 8. This can be conceived of as (more or less) "bad luck" relating to the fact that a system adapting to statistical regularities cannot account for all extreme cases (or "one cannot protect oneself from all eventualities", relating to Moray's, 2003, discussion on eutactic behaviour). An example of this would be when driving straight through an intersection with the right of way, where one normally (and for good reasons) assumes that no other road user will enter the road. An analysis of intersection crashes in naturalistic data reported in Engström et al. (2013) indicated that drivers on a straight path through an intersection with the right of way who crashed with a turning vehicle (which typically did not have the right of way) often did not slow down or visually scan the intersection, which would likely have prevented many of these crashes, even if these drivers were not "at-fault". In terms of predictive processing, this behaviour can be explained in terms of a strong (high-precision) prediction, shaped by extensive previous exposure to similar situations, that no other road users will appear in the intersection.

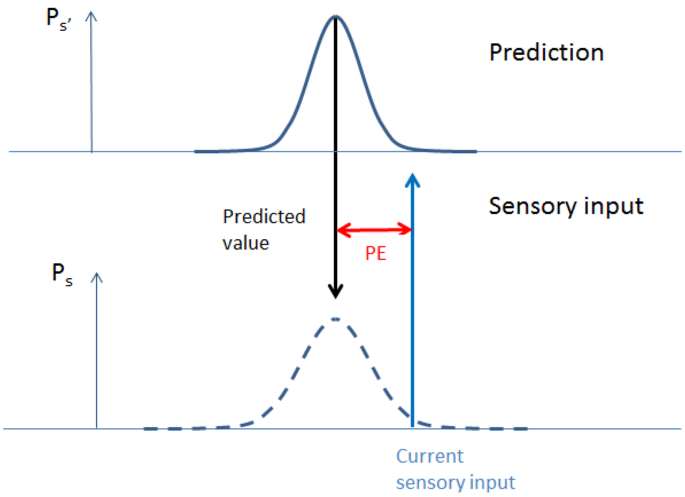

Figure 8: Accurate model but improbable event.

False certainty

This category is characterised by the application of a generative model yielding accurate predictions with high precision, which, however, fail to match an inherently uncertain (i.e., variable, low-precision) environment. Thus, as illustrated in Figure 9, a relatively frequent event (e.g., a lead vehicle braking in dense traffic) would come as a surprise to a driver with a generative model erroneously predicting a low probability of such events. It should be noted that the only difference to the previous category is in the environmental statistics where, in the present case, the precision of the prediction is too high, leaving the driver vulnerable to relatively frequent events (and not just extremely rare events). Thus, in this case it is the *precision estimate* that does not match reality (it is too high), while the average prediction does match how things normally play out. An example of this would be a novice driver following another vehicle deciding to look away from the road to send a text message. Due to her limited experience of such situations, she may judge that it is unlikely that the lead vehicle will brake in this situation while it is in fact relatively likely.

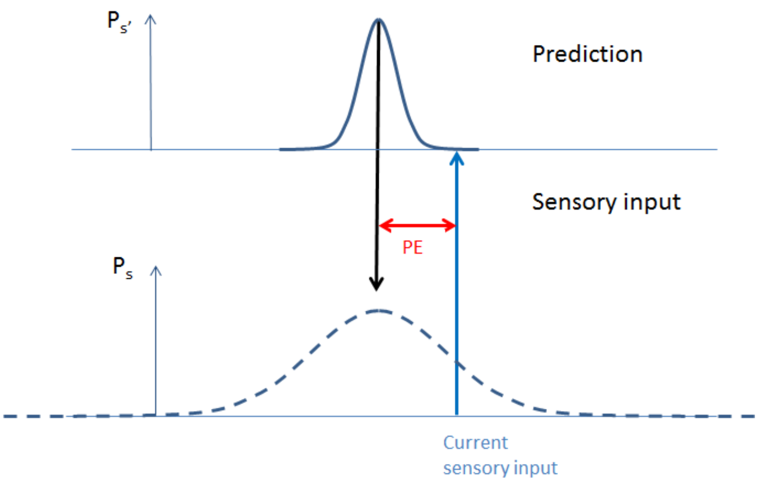

Figure 9: False certainty.

False uncertainty
This type of expectation mismatch, illustrated in Figure 10, represents the opposite to the previous category: the predictions generated by the model have low precision and are thus not trusted by the driver, while the actual traffic environment, and the resulting sensory input, is relatively stable and predictable. This may be the case when driving in a novel, but yet reliable and predicable, environment. This type of situation often leads to a *reduction* in crash risk since uncertain predictions lead to more cautious behaviour, for example in terms of a stronger propensity for epistemic action (e.g., visual scanning) to reduce the (overestimated) uncertainty. A classic example of this is the shift from left-hand to right-hand traffic in Sweden in 1967, which was expected by many to increase crash rate. However, the actual outcome was the opposite, with a significant drop in the number of crashes that only after two years returned to the pre-existing level (Alexandersson, 1972). In terms of the present framework, this could be understood in terms of reduced precision of drivers' predictions leading to more cautious behaviour, thus effectively preventing the three previous types of expectation mismatches (which all involved relatively high-precision predictions). However, with extensive exposure to right-hand traffic, the Swedish drivers' generative models eventually adapted, yielding predictions with higher precision which, again, left drivers more vulnerable to expectation mismatches.

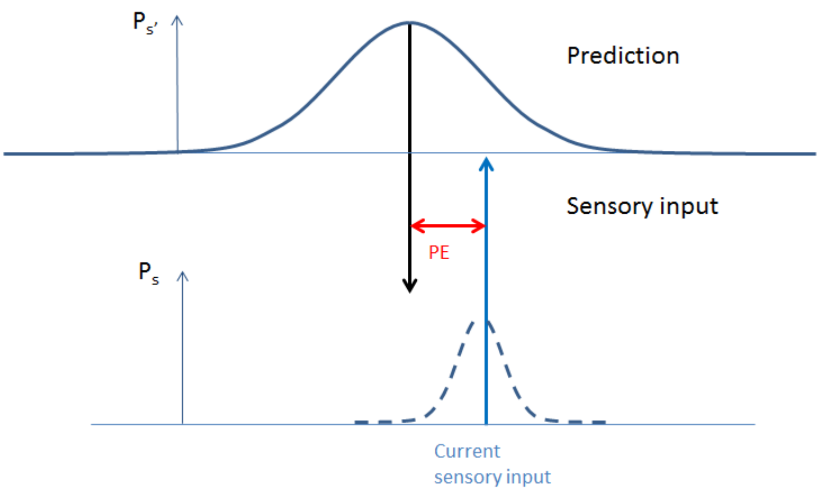

Figure 10: False uncertainty.

Countersteering

Safety functions generating automatic braking and steering interventions in critical situations have a strong potential to prevent crashes. While automatic braking is becoming a standard feature in modern vehicles, a fundamental problem in the design of automatic steering interventions is the phenomenon of *countersteering*. This has been experimentally demonstrated to occur in response to automatic steering interventions (Benderius, 2014; Brockmann et al., 2013), and may be explained as due to activation of the automatic spinal stretch reflex. When the arm muscles are stretched by the steering wheel movement induced by the steering intervention, this is sensed by proprioceptors in the muscle (muscle spindles) which activate alpha neurons in the spinal cord which, in turn, generate a muscle contraction countering the steering intervention (Benderius, 2014). This is clearly an undesired phenomenon since it counteracts the intended purpose of the steering intervention.

The predictive processing framework offers a natural interpretation of this phenomenon which leads to some specific ideas on how it may be mitigated. Recall that action, according to predictive processing, is generated by spinal reflexes automatically cancelling proprioceptive prediction errors. These prediction errors may be understood as the difference between the proprioceptive input (e.g., the muscle state signal from muscle spindle) and descending (top-down) proprioceptive predictions from higher levels in the brain to the spinal cord (traditionally conceived as action commands). When the steering intervention is initiated unexpectedly to the driver, the muscle spindle indicates that the muscle has been stretched while the top-down prediction suggests that it should remain in its original position, thus generating a prediction error that is automatically cancelled by the stretch reflex, producing the countersteering movement.

Based on this, it may be suggested that countersteering may be mitigated if the steering action is cued or gradually ramped up so that the steering intervention is already accommodated by the top-down proprioceptive prediction once it occurs. In this case, the predicted muscle extension will match the actual extension and no prediction error will be generated. Thus, the stretch reflex will be disabled and no countersteering will occur. However, the proposed framework further suggests that this requires frequent exposure to the steering intervention so that the behaviour of the steering intervention is learned, and eventually predicted, by the generative model. This may be accomplished by a system that generates gradual steering interventions also in less critical situations, such as in response to lane deviations (see Abbink et al., 2012, and Abbink, Mulder and Boer, 2012, for existing work along these lines). The predictive processing framework thus leads to quite specific predictions and design suggestions that could be tested experimentally.

Interacting with automated driving functions

Automated driving (AD) is an increasingly important technology development area today in the automotive industry. Some automated functions are already on the market, such as adaptive cruise control with lane centering and automated parking, and others with higher levels of automation (e.g., cars that drive autonomously in pre-defined scenarios or automated truck platoons), are being tested on public roads. A key concept used in contemporary discussion on AD is *levels of automation* (NHTSA, 2013; SAE, 2014), representing increasingly more advanced forms of automation in terms of the degree to which the driver needs to supervise the automation and intervene in situations in which the automation is not able to respond appropriately. For example, in SAE Level 2 automation, the driver is required to participate in the dynamic driving task by monitoring the driving environment and by providing immediate fallback if needed. In Level 3, the driver is not required to monitor the driving environment but is expected to respond appropriately to a request to intervene. In a Level 4 AD function, responsibility for safe operation lies solely with the vehicle, as long as the autonomous driving occurs in certain, pre-defined, scenarios (Seppelt and Victor, 2016).

Key human factors challenges related to AD include the role of the human *during* automated driving, the transfer of control between the AD function and the driver, and the design of human machine interfaces (HMIs) that support system transparency (e.g., communicates system limitations and capabilities; Gibson et al., 2016; Seppelt and Victor, 2016). Here, we explore how predictive processing may provide a novel perspective on these challenges, focusing on three key concepts that are frequently used in discussions about human factors challenges for AD: (1) *drivers' mental models*, (2) *trust in automation* and (3) *being in or out of "the loop"*.

A mental model is a concept commonly used for representing a driver's understanding of an automation function, for example in terms of its different modes of operation and limitations. Mismatches between the driver's mental model and actual system functionality have been suggested as a potential cause of road crashes (Seppelt and Lee, 2015). The predictive processing framework suggests a way to extend the concept of a mental model in terms of a hierarchical generative model generating predictions on how *the actions of an AD function generate sensory input*.

As outlined above, predictive processing suggests that frequent exposure to reliable statistical regularities in the driving environment will lead to a gradual attunement of the generative model and increasingly automatized performance. This idea may be extended to AD; as long as an AD function works reliably and is frequently used, the generative model will eventually produce predictions of sensory input resulting from how the automation controls the vehicle in response to the environmental state. Thus, with learning, the effects of the automation on vehicle control eventually become part of the vehicle-environment statistical regularities that the driver becomes *attuned* to. For example, a driver attuned to Adaptive Cruise Control (ACC) will implicitly predict the visual (e.g., looming) and kinaesthetic (e.g., perceived deceleration) sensory input that is generated by ACC when it slows the vehicle down to maintain a set headway to a slower lead vehicle. This implies that, in case of a system failure (e.g., the ACC fails to slow down), the driver will respond to sensory prediction error (rather than sensory signals per *se*) in a similar fashion as discussed above for braking reactions in critical rear-end situations

These concepts may be useful for understanding drivers' failures in interaction with imperfect automation. As discussed by Seppelt and Lee (2015), problems may occur if the driver's mental model does not properly account for the functional limitations of the automation (e.g., the maximum braking level that the ACC is able to exhibit). The present framework extends this notion by suggesting that such failures may be understood in terms of limited exposure to situations where the automation approaches its functional limitations, thus establishing an inappropriately tuned generative model that fails to account for such situations. Based on this, it may be further suggested that, for the generative model to become *appropriately* attuned to the AD function (i.e., making accurate predictions with a warranted degree of precision; cf. Figures 7-10), the AD function needs to provide transparent feedback on the state of the automation (Bennett & Flach, 2011; Flemisch et al., 2015; Seppelt & Lee, 2007). Such feedback, which may be provided through ecological displays (Vicente and Rasmussen, 1992), then becomes a key part of the statistical regularities of sensory input that the model can tune into. If appropriately designed, the human-machine interface will then support the development of a generative model that properly accounts for the functional limitations of the system, thus correctly predicting what the AD function can and cannot do.

Trust is another key concept in automation research (e.g., Lee and Moray, 1994; Lee and See, 2004). A key issue here concerns the extent to which operators trust the automation more than warranted, what is often referred to as *overreliance*. As reviewed in Lee and See (2004), trust is a complex psychological phenomenon involving both analytical and affective components and potentially influenced by a wide range of individual, organizational, cultural and environmental factors. A complete account of trust is thus clearly beyond the scope of the present paper. Here, we focus on exploring some implications of the idea that trust in automation may be viewed in terms of the *precision* associated with the generative model's predictions regarding (the sensory consequences of) AD function operation[10]. The actions of a reliable automation function system will, with exposure and attunement, be predicted by the generative model with increasingly higher precision (implying increasingly automatized interaction as the driver becomes increasingly attuned to the function). Problems related to overreliance may thus occur if the generative model does not accurately predict actual AD operation (relating to the case of an inaccurate model discussed above in the context of expectancy mismatches, see Figure 7) or if the precision estimates for the model predictions on AD function operation do not match the actual variability in function performance (corresponding to the case of false certainty; see Figure 9). For example, if an active steering function does not work reliably when the road is covered by snow, and this condition is rarely encountered, a driver may predict with an unwarranted degree of precision that the system will handle this condition. Furthermore, even if the predictions and associated precision estimates match actual AD operation, high precision (trust) will increase the vulnerability of the driver to (rare) technical failures (corresponding to the case of "accurate model but improbable event" discussed above, see Figure 8). In this way, one can conceive of different potential "failure modes" of an AD function and how they relate to how the driver's generative model is attuned to the function, with respect to both the accuracy and the precision of model predictions.

A further a key aspect of trust in AD is that a highly trusted automation tends to be monitored less frequently (Lee and See, 2004). Based on the present framework, such monitoring can be viewed as a form of epistemic action. Recall that the role of epistemic (knowledge-seeking) actions such as visual scanning is to increase the precision (reduce uncertainty) associated with predictions for pragmatic actions. In this case, if the precision of the prediction that the AD function will keep the vehicle safely on the road is sufficiently high (i.e., the driver AD function is highly trusted by the driver), the monitoring is no longer needed as it will not contribute to the minimization of prediction error (just like drivers in the intersection scenario above found it unnecessary to scan to the right as hazards typically appeared from the left). In such a situation, the driver may decide to give up monitoring (partially or completely), for example to be able to engage in other tasks, as further discussed below.

Lee and See (2004) emphasizes the affective aspects of trust, suggesting that "…trust also seems to depend on an affective response to the violation or confirmation of implicit expectancies" (p. 61). In terms of predictive processing, such affective responses may be understood in terms of interoceptive predictions of emotional signals

---

[10] This idea resonates with some existing definitions of trust from research on interpersonal relationships, such as that suggested by Rempel et al., 1985, quoted in Lee and See (2004, p. 54): " expectations related to the subjective probability an individual assigns to the occurrence of some set of future events".

(see Pezzulo, 2013; Seth, 2013) associated with exteroceptive predictions of external stimuli, such as looming, generated by the AD function. Thus, such *embodied inference* (Pezzulo, 2013) seems to be an key component in a more comprehensive predictive processing-based account of trust in automation.

A third concept, abundant in discussions on AD, is being “in or out of the loop” (Endsley and Kiris, 1995; Seppelt and Victor, 2016). For example, being in the loop during automated driving generally refers to monitoring the driving environment and being aware of the status of the automation. Conversely, driver *out-of-the-loop unfamiliarity* (OOTLUF) has been defined as a decreased ability to detect system errors and to intervene and perform the task(s) manually in response to failures (Seppelt and Victor, 2016). However, a detailed, mechanistic, account of “the loop” and what it means to be “in” or “out” of it is currently lacking. Such an account is sketched below based on the PP framework.

In predictive processing terms, being in the loop for a driving task can be defined as *being engaged in inference related to that task*. Being engaged in inference means continuously generating sensory predictions and cancelling prediction errors by means of perception and/or action. If prediction errors are not only minimized by updating predictions but also by actively cancelled by action (e.g., steering and braking) this corresponds to *active inference*. Importantly, as discussed above, such inference can occur at different levels of the driving task (e.g., operational, tactical and strategic; Michon, 1985). Driving manually without AD usually entails being in the loop at all three levels, where the higher levels generate top-down predictions for the level below, as discussed above. However, driving with AD alters this picture. In Level 2 automation, as mentioned above, the vehicle performs steering and speed/headway control while the driver is still expected to continuously monitor the situation and quickly resume control if required. In this situation, the driver is no longer engaged in active inference but merely in purely perceptual inference, similar to a passenger that monitors the road ahead and the operation of the driver. At the operational level, the generative model may thus make predictions of how sensory signals (e.g., looming) will change due to the operation of the automation. If so, the driver is still engaged in inference at the operational level, and thus, according to the proposed definition, still in the loop at that level. However, this type of passive inference could be expected to be qualitatively different from the full active inference in operation during manual driving (whether this is actually the case is an interesting empirical question).

Alternatively, the driver may choose to just monitor the general situation, in which case she is out of the operational loop but remaining in the tactical and strategic control loops. As discussed above, the monitoring required during Level 2 driving can be viewed as a form of epistemic action with the purpose to reduce uncertainty of predictions (in this case related to the drivers’ trust in the AD function). Thus, if the driver does not fully trust the automation (or the passenger don’t fully trust the driver), she will perceive a need for monitoring to reduce uncertainty. Conversely, if the driver has a high degree of confidence in the AD system she may feel less inclined to monitor its operation. At very high levels of trust, the driver may find the need for monitoring at the operational level unnecessary and thus only engage in more cursory monitoring at the tactical level, or even engage in secondary tasks such as reading a newspaper. In the latter case, the driver is no longer engaged in inference at the operational and tactical levels and can be regarded as being out of the loop at these levels, left only in the strategic loop[11]. This is clearly a great safety concern given the limited ability of Level 2 AD systems to handle critical situations.

In Level 3 automation, the driver may still be engaged in monitoring, but there is no requirement to do so. Thus, inference at the operational or tactical levels are at best very limited (again depending on the degree of trust in the function). This may explain why it takes drivers surprisingly long to resume control from Level 3 automation mode (e.g., 5-7 seconds, as suggested by Gold et al., 2013); when requested to take over control when being out of the operational and tactical loops, the driver first needs to re-establish predictions at the operational and tactical level in order to generate prediction errors that can be acted upon.

Relation to existing human factors theories and models

In this section, we review some of the main theoretical approaches in contemporary human factors research and discuss how they relate to the predictive processing framework.

Mental models, schemata and the abstraction hierarchy

As already mentioned, a common approach for conceptualizing driver knowledge (e.g., of driving support and automation functions) and expectancy in driving is in terms of mental models (e.g., Rouse and Morris, 1985; Seppelt and Lee, 2015). As discussed in the context of automated driving functions, the concept of a hierarchical generative model may be viewed as a specific form of mental model, representing how sensory input is generated by states (or "causes") in the external world and the own body. A closely related construct is *schemata* (e.g., Arbib, 1995; Engström et al., 2013; Theeuwes and Godthelp, 1995). A schema may refer to basic motor actions such as "brake", routine action sequences such as "overtake" or more general tasks such as "drive home", and are often viewed in terms of long-term memory structures established through experience and training (Endsley, 1995). Schemata are typically viewed as hierarchically organized, where higher-level schemata (e.g., "turn right at T-junction") provide context for the lower-level schemata (e.g., "look-left"). Like mental models, schemata are hence closely related to the concept of a hierarchical generative model, in which individual sub-schemata (e.g., "brake" or "drive home") would correspond to predictions at different levels.

A related framework is the *abstraction hierarchy* proposed by Rasmussen (1985) which represents goal-relevant constraints operating in a work domain at different levels of abstraction. Higher levels represent more abstracted, lower resolution properties of a domain as compared to the lower, more concrete levels. The levels are linked by structural means-ends relations, spanning the gap from functional purpose (at the highest level) to material form (at the lowest-level). In order to successfully control a process in a specific domain, the operator must embody a hierarchical model of the constraints operating at the different levels (Vicente and Rasmussen, 1992). Again, this

[11] However, even when engaged in a secondary task, the driver may still to some extent be engaged in passive inference in other sensory modalities, for example, by continuously predicting somatosensory and proprioceptive input.

resonates well with the predictive processing concept of a hierarchical generative model representing how external states or causes (corresponding to constraints in Rasmussen's terms) produce sensory input. In particular, predictive processing offers a specific mechanistic proposal for the cognitive mechanisms governing the inference of causal relationships represented in the abstraction hierarchy.

What mainly distinguishes the present account from these more traditional hierarchical frameworks is the fundamental role assigned to prediction and the single unifying principle of prediction error minimization. In particular, viewing schemata or means-end relations in the abstraction hierarchy in terms of predictions leverages the concepts of inference, precision and model tuning which, as we have seen, yield new insights into a range of different driver behaviour phenomena. Predictive processing further offers a solid mathematical foundation for hierarchical modelling and is, by contrast to most traditional human factors models, firmly based in contemporary neuroscience.

Automaticity and the skills, rules, knowledge (SRK) taxonomy

We have already discussed how the predictive processing framework may offer an account of learning and the development of automaticity in terms of a gradual tuning of the generative model to statistical structures in the environment. This is closely related to the *skills, rules, knowledge (SRK) taxonomy* proposed by Rasmussen (1983). Like predictive processing, this framework suggests that, with increasing experience, operators become *attuned* to perceptual features of the environment, leading to a shift from (controlled) knowledge-based to (automatized) skill and rule-based performance. This leads to the concept of *ecological interface design*, emphasizing the reciprocity between the operator and the work environment (Vicente and Rasmussen, 1992). The present framework seems fully in line with these concepts and additionally offers a concrete mechanistic account for *how* operators become attuned to their work environment through gradual adjustment of generative model parameters driven by the goal to minimize prediction error. A concrete example of this, discussed above, is drivers' attunement to automated driving functions and how this may be supported by ecological interfaces properly conveying their functional limitations.

Situation awareness

Another very influential theoretical construct in human factors is situation awareness (SA). SA was developed to address how an operator (or a team of operators) functions in a dynamical environment and is typically described in terms of three levels or phases: (1) perceiving and attending to cues in the situation, (2) comprehending the meaning of the perceived information and (3) projecting the future state of the situation (Endsley, 1995).

SA can thus be seen as representing the current state of knowledge or, alternatively, as a mental model of the current situation (Endsley, 1995). This model is enabled by various mental processes and knowledge structures in long-term memory, where the latter are traditionally conceived of in terms of schemata established through experience and training, as discussed above.

The SA account outlined by Endsley (1995, 2015) has several key ideas in common with the present framework. In particular, the proposed mechanism for updating the situation model (SA) corresponds closely to (perceptual) inference in predictive processing:

> "The main clue to erroneous SA will occur when a person perceives some new piece of data that does not fit with expectations based on his or her internal model. When a person's expectations do not match with what is perceived, this conflict can be resolved by adopting a new model, revising the existing model, or changing one's goals and plans to accommodate the new situation classification. If the new data can be incorporated into the model, this may merely indicate that a new prototypical situation (state of the model) is present that calls up different goals and plans accordingly. If the new data cannot easily fit into the existing model, the model may be revised. A common problem is whether to continue to revise the existing model to account for the new data or choose an alternate model that is more appropriate. For the latter to occur, something about the data must flag that a different situation is present." (Endsley, 1995, p. 57)

Moreover, Endsley (1995, p. 44), based on Holland et al. (1986), suggests a mechanism for the development of mental models which directly parallels the "bootstrapping" account of learning and model tuning outlined above (and in Clark, 2016), where the generative model parameters are gradually tuned to minimise prediction errors. Endsley's framework also addresses the *confidence level* of SA (1995, p. 45), which is closely related to the concept of precision in predictive processing.

However, while the traditional SA framework relies heavily on various traditional information processing concepts deriving from the computer metaphor (storage, limited capacity, pointers in memory etc.), predictive processing potentially offers a more detailed, biologically grounded account of the cognitive mechanisms underlying the key SA concepts. Furthermore, while prediction plays a key role in SA, it comprises only one out of three key elements in the theory (the other two being perception and comprehension). By contrast, in the present framework prediction takes centre stage with prediction error minimization as the single unifying principle based on which all other cognitive functions can be understood. Hence, "perception" and "information integration" *emerges* as a result of the prediction and the resulting prediction error minimization.

Finally, while SA emphasizes the role of action, the inference mechanism outlined in the quote above from Endsley (1995) focuses on passive perceptual inference. The *active inference* concept in the present framework extends this notion to action, thus offering a more complete framework for how perception and action are interrelated when an operator interacts with a dynamic environment. This also potentially enables a closer contact between the SA concept and action-oriented approaches to cognition such as active vision (Hayhoe and Ballard, 1995), ecological psychology (Gibson, 1979) and embodied cognitive science (Clark, 2013; 2016).

Predictive processing thus aligns closely with the traditional SA framework at the general level and several of the general concepts presented here were already addressed in Endsley's (1995) seminal paper. However, the predictive processing framework potentially extends the traditional SA framework in several ways. These include, in particular, the view of prediction as primary, the unifying role of prediction error minimization and the role of action in fulfilling predictions (active inference). These concepts may potentially offer a more precise and biologically grounded foundation for SA and its underlying mechanisms.

The SEEV model

SEEV (Salience, Effort, Expectancy, Value) was originally developed for modelling visual sampling patterns in aviation (Wickens et al., 2003) but is also frequently used in the automotive domain (e.g., Horrey, Wickens and Consalus, 2006). In SEEV, expectancy and value are viewed as top-down knowledge-driven factors while salience and effort are considered bottom-up factors. Expectancy is defined in terms of information content or, more precisely, information *bandwidth*. The bandwidth of a "channel" (e.g., an area of interest in a visual scene) represents the amount of information per time unit (bits/s) in that location. A location with a high rate of information is expected to be sampled frequently, as empirically supported by the early visual occlusion studies by Senders (1963, 1980; see also Kujala et al., 2016). Information sampling is also assumed to be driven top-down by value, which represents the perceived *utility* of sampling certain information. The probability that information will be observed at a particular location and the value of this information combine in an *expected value* model of scanning, in which both high event frequency and high value regions drive sampling rates.

The basic ideas behind the SEEV model bear a close resemblance to the predictive processing account of visual sampling (Friston et al., 2012) which, as described above, is also based on information theoretic principles. In particular, the idea that operators sample the environment based on expected bandwidth is directly paralleled by the present proposal that the key role of visual sampling (or epistemic action in general) is to minimize uncertainty, thus minimizing overall prediction error. Indeed, in information theoretic terms, prediction error corresponds to *surprise,* which is equivalent to (self-) information (the negative log probability of an outcome; Friston, 2010). From the general mandate to minimize prediction error, it follows that areas should be sampled in proportion to where the expected surprise (i.e., information content) is largest, which is also the key notion underlying SEEV. The predictive processing framework further suggests that valuable information is simply that which minimizes surprise, thus suggesting a unified view of expectation and value that seems to sit comfortably with the SEEV notion of expected value.

What predictive processing offers beyond SEEV is a broader theory for the how visual sampling strategies emerge from the single imperative to minimize prediction error. While outside the scope of the present paper, a more formal comparison between the predictive processing account of visual sampling, SEEV, the related classical work by Senders (1964, 1983) and more recent accounts such as Kujala et al. (2016) is clearly an interesting topic for future exploration.

Motivational models

A strong tradition in applied driving research has focused on understanding adaptive driver behaviour, in particular behavioural adaptation to safety interventions. This has involved the development of a number of influential driver behaviour models, which are often referred to as *motivational* (Ranney, 1994). Motivational driver models typically have their roots in cybernetics and control theory, and generally suggest that drivers regulate their behaviour (e.g., selected speed) against some criterion, or reference value (although the nature of this reference value has been hotly debated). Examples include the risk homeostasis theory (Wilde, 1982), the zero-risk theory (Summala, 1988), the threat avoidance model (Fuller, 1984) and the Extended Control Model (Hollnagel, Nåbo and Lau, 2003; see Engström and Hollnagel, 2007, for a review). More recently, there has been a convergence towards the idea that drivers regulate behaviour based on emotions and feelings, building on Damasio's (1994) concept of *somatic markers* (e.g., Fuller, 2007 Summala, 2007; Vaa, 2007). More specifically, Summala (2007) suggested that drivers regulate behaviour in order to avoid feelings of *discomfort*, thus attempting to remain in a subjectively defined *comfort zone* which defines the minimum accepted *safety margins* adopted (Engström and Ljung Aust, 2011; Ljung Aust and Engström, 2011).

While only briefly touched upon above, the predictive processing framework offers a range of concepts that could extend these notions toward a more detailed, mechanistic account of how driver behaviour is governed by emotions and feelings. In particular, this includes the notion that the generative model not only producing predictions in the exteroceptive and proprioceptive domains, but also interoceptive predictions on expected emotional responses (see Figure 1 and Figure 6). In the predictive processing literature, models of interoceptive inference has been developed in particular by Pezzulo (2013) and Seth (2013) (see also Pezzulo et al. 2015). This leads to the idea of *embodied inference* (Pezzulo, 2013), suggesting that exteroceptive (e.g., looming) predictions become correlated with interoceptive, emotional predictions (e.g. relating to discomfort). Applying these ideas to driving may offer a novel perspective on the nature and origin of drivers' comfort zones (Summala, 2007), adaptive driver behaviour as well as affective aspects of trust, as briefly discussed above. This clearly constitutes a very interesting avenue for further research.

Ecological psychology and visual guidance

Finally, predictive processing also fits well with accounts of visually-guided vehicle control deriving from ecological psychology (e.g., Gibson, 1979). There is a long-standing research tradition studying how humans and other animals use optical invariants to guide behaviour, including several applications to driving. For example, Lee, (1976) showed that controlling braking by keeping the optical variable tau-dot at a constant value of -0.5 will lead to a smooth stop just in front of an obstacle. Similarly, Land and Lee (1994) showed that steering in a way that keeps the perceived tangent point at a constant visual angle will generate a smooth path around a bend and (as mentioned above) Wann and Land (2000) showed that, in the absence of a tangent point, a smooth turn towards a desired target is generated by fixating the target and keeping speed and keeping the visual angular rate constant.

These types of models dovetail nicely with the present concept of active inference, where the key role of action is to actively align sensory input with current predictions, thus minimizing prediction errors. Thus, in terms of predictive processing, the use of perceptual invariants to guide control amounts to predicting that a certain perceptual quantity will remain constant and then acting (e.g., by steering or braking) to minimize deviations from this prediction.

Thus, again, the present account seems fully in line with ecological psychology models, but additionally offers a framework for unifying ecological models of perception and action with other, traditionally disparate, cognitive phenomena such as attention and learning, based on the fundamental unifying principle of prediction error minimization.

Conclusions

This paper offered an initial exploration of how predictive processing concepts could be applied in the domain of automobile driving. A central general theme that emerges from the predictive processing account of driving is how drivers' expectations are shaped gradually by the tuning of an internal generative model to statistical regularities of the environment. Such statistical regularities may, for example, include typical movement patterns of other road users in a certain type of scenario (e.g., "hazards typically appear to the left") or the operation of driving support and automated driving functions (e.g., "the active steering function reliably keeps the vehicle in the lane"). In particular, the present account suggests that expectancy "failures" leading to crashes and overreliance on automation can both be understood in terms of inappropriate tuning of the generative model leading to inaccurate and/or overconfident predictions. This further suggests that a key goal in the design of vehicle systems and road infrastructure should be to support predictions that are both accurate and associated with precision estimates that match the actual variability of the operating environment (i.e., yield predictions with an appropriate level of confidence). According to predictive processing, preventing too high precision estimates (false certainty) will promote epistemic actions (e.g., increased visual scanning, slowing down to increase the time available to sample the environment or, in the case of AD, increased monitoring of the road environment) to control uncertainty, resulting in more cautious driving behaviour.

Another general implication is that, at least for experienced, well-attuned drivers, driving behaviour can, to a large extent, be predicted from the statistical environmental regularities that shaped the behaviour. Concrete examples of this include the finding that driver reactions in emergency situations is largely determined by situation kinematics (Markkula et al., 2016) and the observation of stereotypical scanning patterns in intersections, governed by statistically frequent configurations of traffic elements (Räsänen and Summala, 1998).

There are also several other potential application areas that could not be covered here due to space limitations. This includes the understanding of driver distraction and inattention (Engström, Monk et al., 2013; Lee, Regan and Young, 2008) as well as the development of individual driving styles and behavioural change programs intended to promote safer and more efficient driving behaviour (Sagberg et al., 2015).

As reviewed above, several of the general concepts underlying the present framework, such as hierarchical mental models, the interaction between top-down and bottom-up processing, the key role of expectation, confidence in expectations and reinforcement learning are accounted for by existing human factors frameworks and models such as the Abstraction Hierarchy, Situation Awareness and the SEEV model. Thus, predictive processing should not be viewed as a radical alternative to these existing models but rather as a framework for bringing together different strands of human factors research based on the unifying principle of prediction error minimization. Due to the generality of this basic underlying principle, predictive processing has been proposed as a unifying framework for basic cognitive science (Clark, 2013, 2016) and neuroscience (Friston, 2010; Huang, 2008) as well as psychiatry (Friston et al., 2014) and seems possible that it could also play a similar future role in applied human factors and the study of driver behaviour. In particular, predictive processing may offer a more precise characterization of traditional human factors concepts such as (hierarchical) mental models, situation awareness, expectancy and learning, supported by a solid mathematical framework and grounded in contemporary neuroscience.

In addition, the predictive processing framework offers a set of specific concepts which can be regarded as novel contributions to human factors research. These include in particular:

- The extended view of a mental model as a generative model of how sensory input is produced by external or bodily states
- The view of action as the minimization of prediction errors by aligning sensory input with the prediction (i.e., "bringing about" the predicted input)
- The idea that prediction/expectancy is primary and always present ("there is always a prediction of something")
- The concept of precision and the versatile ways in which it can be used to control behaviour
- The idea of embodied inference (Pezzulo, 2013) enabled by correlations between exteroceptive and interceptive (emotional) predictions

The present paper was intended as a first exploration of the application of predictive processing to driving and we hope that it will encourage others to apply and further develop these ideas. Efforts towards more specific quantitative driver behaviour models based on predictive processing concepts are currently underway and will be reported in future publications.

References

Abbink, D.A., D., Cleij, M., Mulder and M.M van Paassen. 2012. The Importance of Including Knowledge of Neuromuscular Behaviour in Haptic Shared Control. Paper presented at the IEEE International Conference on Systems, Man, and Cybernetics (SMC). Seoul, South Korea.

Abbink, D., M. Mulder, and E. Boer, E. 2012. Haptic Shared Control: Smoothly Shifting Control Authority? *Cognition, Technology & Work* 14(1):19-28. doi: 10.1007/s10111-011-0192-5

Adams, R. A., S. Shipp, and K. J. Friston. 2013. Predictions Not Commands: Active Inference in the Motor System. *Brain Structure and Function* 218 (3): 611-643. doi: 10.1007/s00429-012-0475-5

Alexandersson, S. 1972. *Some Data About Traffic and Traffic Accidents*. The Swedish Road Safety Office, Stockholm.

Allen, T.M., H. Lunenfeld and G. J. Alexander. 1971. Driver Information Needs. *Highway Research Record* 366: 102–115.

Arbib, M.A. 1995. Schema theory. In MA Arbib (Ed.), *Handbook of brain theory and neural networks*. 830-834. Cambridge: Massachusetts Institute of Technology.

Ashby F. G., B.O Turner and J. C Horvitz. 2010. Cortical and Basal Ganglia Contributions to Habit Learning and Automaticity. *Trends in Cognitive Science*, 14(5):208-215. doi: 10.1016/j.tics.2010.02.001

Aston-Jones, G. and J. D Cohen. 2005. An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Reviews of Neuroscience*, 28:403-450. doi: 10.1146/annurev.neuro.28.061604.135709

Benderius, O. 2014. *Modelling driver steering and neuromuscular behavior*. PhD Thesis, Chalmers University, Gothenburg, Sweden

Badre, D. 2008. Cognitive Control, Hierarchy, and the Rostro-caudal Axis of the Frontal Lobes. *Trends in Cognitive Science*, 12(5), 193-200. doi: 10.1016/j.tics.2008.02.004

Bar, M. 2007. The Proactive Brain: Using Analogies and Associations to Generate Predictions. *Trends in Cognitive Sciences* 11(7): 280–89. doi: 10.1016/j.tics.2007.05.005.

Bennett, K. B. and J. M. Flach. 2011. *Display and Interface Design: Subtle Science, Exact Art*. Boca Raton: CRC Press.

Braver T. S. and J. D. Cohen. 2000. On the Control of Control: The Role of Dopamine in Regulating Prefrontal Function and Working Memory. In S. Monsell and J. Driver (Eds.) *Control of Cognitive Processes, Attention and Performance XVIII*. doi: 10.1002/acp.849

Boer, E. R. 1999. Car Following From the Driver's Perspective. *Transportation Research Part F: Traffic Psychology and Behavior, 2*(4): 201–206. doi: 10.1016/S1369-8478(00)00007-3

Brockmann, M., E. Johansson, A. Rambaldini, T. Hesse, F. Nawrat et al. 2013. *InteractIVe Deliverable D3.1, Results from IWI Evaluation. Executive Summary*. http://www.interactive-ip.eu/index.dhtml/docs/interactIVe-D3.1-ResultsfromIWIEvaluation-RE-ExecutiveSummary.pdf

Brown, H., R. A. Adams, I. Parees, M. Edwards and K. Friston. 2013. Active Inference, Sensory Attenuation and Illusions. *Cognitive Processing*, 14 (4): 411-427. doi: 10.1007/s10339-013-0571-3

Clark, A. 2016. *Surfing uncertainty*. Oxford: Oxford University Press.

Clark, A. 2013. Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences*, 36 (3): 181-204. doi: 10.1017/S0140525X12000477

Cohen, J.D., K. Dunbar and J. L. McClelland. 1990. On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. *Psychological Review*, 97(3): 332-361. doi: 10.1037/0033-295X.97.3.332

Cooper, R. and T. Shallice. 2000. Contention Scheduling and the Control of Routine Activities. *Cognitive Neuropsychology* 17(4): 297 - 338. doi: 10.1080/026432900380427

Deco, G. and E. T. Rolls. 2005. Attention, Short-term Memory and Action Selection: A Unifying Theory. *Progress in Neurobiology* 76(4): 236-256: doi: 10.1016/j.pneurobio.2005.08.004

Damasio, A. R. 1994 *Descarte's error: Emotion, Reason and the Human Brain*. New York: Groset/Putnam.

den Ouden, H.E.M., P. Kok, and F.P de Lange. 2012. How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology* 3(548). doi: 10.3389/fpsyg.2012.00548

Endsley, M. R. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1): 32–64. doi: 10.1518/001872095779049543.

Endsley and Garland, 2000 (Eds.). *Situation Awareness Analysis and Measurement.* Mahwah, NJ.Lawrence Erlbaum Associates.

Endsley, M. R., and Kiris, E. O. 1995. The out-of-the-loop performance problem and level of control in automation. Human Factors, 37(2), 381-394.

Engström, J. 2011. Understanding attention selection in driving: From limited capacity to adaptive behaviour. PhD Thesis, Chalmers University, Sweden.

Engström, J., J. Bärgman and J. Lodin. 2016. *EFrame – Evaluation Framework for Commercial Vehicle Safety Systems and Services: Task report of work achieved in WP 4 Task 4.3 (EUC2-4): Safety benefit estimation for an Advanced Emergency Braking System*. EFrame internal project report.

Engström, J. and Hollnagel, E. 2007. A General Conceptual Framework for Modelling Behavioural Effects of Driver Support Functions. In P.C. Cacciabue (Ed.), *Modelling Driver Behaviour in Automotive Environments Critical Issues in Driver Interactions with Intelligent Transport Systems*. London: Springer.

Engström, J., T. Victor and G. Markkula. 2013. Attention Selection and Multitasking in Everyday Driving: A Conceptual Model. In M.A Regan, T.W. Victor and J.D. Lee (Eds.) *Driver Distraction and Inattention: Advances in Research and Countermeasures*. Farnham: Ashgate.

Engström, J., G. Markkula, T. Victor and N. Merat. in press. Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. To appear in *Human Factors*.

Engström, J., C. A. Monk, R. J. Hanowski, W. J. Horrey, J. D. Lee, D. V. McGehee, M. Regan et al. 2013. *A Conceptual Framework and Taxonomy for Understanding and Categorizing Driver Inattention*. US-EU ITS Cooperation, Driver Distraction and HMI Working Group.

Engström, J., J. Werneke, J. Bärgman, N. Nguyen and B. Cook. 2013. Analysis of the Role of Inattention in Road Crashes Based on Naturalistic On-board Safety Monitoring Data. In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*, Gothenburg, Sweden.

Feldman, H. and K. J. Friston. 2010. Attention, Uncertainty, and Free-energy. *Frontiers in Human Neuroscience* 4(215). doi: 10.3389/fnhum.2010.00215

FitzGerald, T. H. B., R. J. Dolan and K. J. Friston. 2014. Model averaging, optimal inference, and habit formation. *Frontiers in Human Neuroscience* 8(457). doi: org/10.3389/fnhum.2014.00457

Flemisch, F. O., K. H. Bengler, H. Bubb, H. Winner and R. Bruder, R. 2015. Towards Cooperative Guidance and Control of Highly Automated Vehicles – H-Mode and Conduct-by-Wire. *Ergonomics*, 57(3): 343-360. doi: 10.1080/00140139.2013.869355

Fletcher, P. and C. Frith. 2009. Perceiving is Believing: A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia. *Nature Reviews: Neuroscience* 10:48–58. doi: 10.1038/nrn2536

Friston, K. J. 2005. A Theory of Cortical Responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 360(1456):815–36. doi. 10.1098/rstb.2005.1622

Friston, K. J. 2008. Hierarchical Models in the Brain. *PLoS Computational Biology* 4(11). doi. 10.1371/journal.pcbi

Friston, K. J. 2010 The Free-energy Principle: A Unified Brain Theory? *Nature Reviews Neuroscience* 11(2):127–38. doi. 10.1038/nrn2787

Friston, K., R. A. Adams, L. Perrinet and M. Breakspear. 2012. Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology*, 3(151). doi. 10.3389/fpsyg.2012.00151

Friston, K. J., J. Daunizeau, J. Kilne and S. J. Kiebel. 2010. Action and Behavior: A Free-energy Formulation. *Biological Cybernetics* 102(3): 227–60. doi. 10.1007/s00422-010-0364-z

Friston, K. J., D. O. Rigoli, C. Mathys, T. Fitzgerald and G. Pezzulo. 2015. Active Inference and Epistemic value. *Cognitive Neuroscience*, 6(4): 187-214. doi. 10.1080/17588928.2015.1020053.

Friston K. J. K. E. Stephan, P. R. Montague and R. J Dolan. 2014. Computational Psychiatry: The Brain as a Phantastic Organ. *The Lancet Psychiatry* 1(2), 148-158

Friston, K. J, T. FitzGerald, F. Rigoli, P. Schwartenbeck and G. Pezzulo. 2016. Active Inference: A Process Theory. *Neural Computation*, Nov 29: 1-49.

Fuller, R. 2007. Motivational determinants of control in the driving task. In P. C. Cacciabue (Ed.), *Modeling Driver Behavior in Automotive Environments: Critical issues in Driver Interactions With Intelligent Transport Systems* (pp. 165–188). London: Springer.

Fuller, R.C.G. 2005. Towards a General Theory of Driver Behaviour. *Accident Analysis and Prevention*, 37: 461–472.

Fuller, R.G.C. 1984. A Conceptualisation of Driving Behaviour as Threat Avoidance. *Ergonomics* 27(11): 1139–1155.

Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston.

Gibson, M., J. D, V. Venkatraman, M. Price, M. and J. Lewis. 2016. Situation Awareness, Scenarios, and Secondary Tasks: Measuring Driver Performance and Safety Margins in Highly Automated Vehicles. *SAE International Journal of Passenger Cars- Mechanical Systems – Electronic and Electrical Systems.* 9(1): 237-242.

Gold C., D. Dambock, L. Lorenz and K. Bengler. 2013. Take over!"—How Long Does It Take to Get the Driver Back Into the Loop? In: *Proceedings of the Human Factors and Ergonomics Society*, 57th annual meeting, 1938–1942.

Gold, J. N. and M. N. Shadlen. 2001. Neural Computations That Underlie Decisions About Sensory Stimuli. *Trends in Cognitive Sciences* 5(10):16 238–55. doi: org/10.1016/S1364-6613(00)01567-9

Goodrich, M. A., W. C. Stirling and E. R. Boer. 2000. Satisficing Revisited. *Minds and Machines,* 10(1): 79–109. doi. 10.1023/A:1008325423033

Green, M., 2000. How Long Does It Take to Stop? Methodological analysis of driver perception-brake times. *Transportation Human Factors* 2(3): 195–216. Doi. 10.1207/STHF0203_1.

Hayhoe, M., & Ballard, D. 2005. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9, 188–194.

Hebb, D. O. 1949. *The Organization of Behaviour*. New York: Wiley.

Helmholtz, H. von. 1876. *Handbuch der Physiologischen Optik*. Leopold Voss.

Hohwy, J. A. 2013. *The Predictive Mind*. New York: Oxford University Press.

Hohwy, J. A. Roepstorff and K. Friston 2008. Predictive Coding Explains Binocular Rivalry: An Epistemological Review. *Cognition* 108(3):687–701. doi. 10.1016/j

Holland, J. H., Holyoak, K.. J., Nisbett, R. E., &Thagard, P. R. 1986. *Induction: Processes of inference, learning and discovery*. Cambridge, MA: Bradford Books/MIT Press.

Hollnagel, E., Nåbo, A. and Lau, I. (2003). A systemic model for driver-in-control. *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*., Park City, UT. Public Policy Center, University of Iowa.
Huang, G. 2008. Is This a Unified Theory of the Brain? *New Scientist* 2658: 30–33
Knill, D. and A. Pouget. 2004. The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation. *Trends in Neuroscience* 27(12): 712–19. doi. 10.1016/j.tins.2004.10.007
Koustanaï, A., E. Boloix, P. van Elslande and C. Bastien .2008. Statistical Analysis of "Looked-But-Failed-To-See" Accidents: Highlighting the Involvement of Two Distinct Mechanisms. *Accident Analysis & Prevention, 40* (2): 461-469. doi: 10.1016/j.aap.2007.08.001
Kujala, T., J. Mäkelä, I. Kotilainen and T. Tokkonen. 2016. The Attentional Demand of Automobile Driving Revisited: Occlusion Distance as a Function of Task-Relevant Event Density in Realistic Driving Scenarios. *Human Factors* 58: 163-180. doi: 10.1177/0018720815595901
Land, M. F. and D. N Lee. 1994. Where We Look When We Steer, *Nature* 369: 742-744. doi: 10.1038/369742a0
Lee, J. D., & K. A, See. 2004. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.

Lee J.D. and N. Moray. 1994. Trust, Self-confidence, and Operators' Adaptation to Automation. *International Journal of Human Computer Studies* 40(1):153–184.
Lee, J.D., M. A. Regan and K. L. Young. 2008. What Drives Distraction? Driver Distraction as a Breakdown of Multi-level Control. In Regan, M., Lee, J. and Young, K. (Eds) *Driver Distraction: Theory, Effects and Mitigation*. Boca Raton: CRC Press.
Markkula, G., 2014. Modeling Driver Control Behavior in Both Routine and Near-Accident Driving. Chicago, IL In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58:879–883.
Markkula, G. 2015. Answering Questions About Consciousness by Modeling Perception as Covert Behavior. *Frontiers in Psychology*, 6. doi. org/10.3389/fpsyg.2015.00803
Markkula, G., J. Engström, J. Lodin, J., Bärgman, and T, Victor. 2016. A Farewell to Brake Reaction Times? Kinematics-dependent Brake Response in Naturalistic Rear-end Emergencies. *Accident Analysis and Prevention* 95: 209–226. doi. 10.1016/j.aap.2016.07.007
Michon, J. A. 1985. A Critical Review of Driver Behaviour Models: What Do We Know? What Should We Do? In Evans L.A. & Schwing R.C. (Eds.). *Human Behaviour and Traffic Safety*. 487–525. New York: Plenum Press.
Miller, E. K. and J. D. Cohen. 2001. An Integrative Theory of Prefrontal Cortex Function. *Annual Reviews of Neuroscience*, 24: 167-202. doi. 10.1146/annurev.neuro.24.1.167
Montague, P.R., S. E. Hyman and J. D. Cohen. 2004. Computational Roles for Dopamine in Behavioural Control. *Nature* 431: 760–767. doi. 10.1038/nature03015
Moray, N. 2003. Monitoring, Complacency, Scepticism and Eutectic Behaviour. *International Journal of Industrial Ergonomics,* 31(3): 175-178.
NHTSA. 2013. US Department of Transportation Policy on Automated Vehicle Development, p 42.

Olson, P.L., 1989. *Driver Perception Response Time*. Technical Report 890731. Society of Automotive Engineers.
Olson, P.L. and M. Sivak. 1986. Perception-response Time to Unexpected Roadway Hazards. *Human Factors* 28 (1): 91–96. doi: 10.1177/001872088602800110
Pezzulo, G., 2013. Why Do You Fear the Bogeyman? An embodied predictive coding model of perceptual inference. *Cognitive and Affective Behavioral Neuroscience*. 14(3): 902-911. doi: 10.3758/s13415-013-0227-x
Pezzulo, G., F. Rigoli and K. J. Friston. 2015. Active Inference, Homeostatic Regulation and Adaptive Behavioural Control. *Progress in Neurobiology*, 134: 17-35. doi:10.1016/j.pneurobio.2015.09.001.
Rao, R. P. N. and D. H. Ballard.1999. Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-classical Receptive-field Effects. *Nature Neuroscience* 2(1):79–87. doi: 10.1038/4580.
Rasmussen, J. 1985. The role of hierarchical knowledge representation in decision making and system management. *IEEE Transactions of System, Man and Cybernetics.* vol. SMC-15. pp. 234-243.
Rescorla, R. A. and A. R. Wagner. 1972. A theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Non-reinforcement. In A. H. Black and W. F Prokasy (Eds.) *Classical Conditioning II: Current Research and Theory*, 64–99. New York: Appleton Century Crofts
Rempel, J. K., J.G., Holmes and M.P Zanna. 1985. Trust in close relationships. *Journal of Personality and Social Psychology, 49*(1), 95–112.
Räsänen, M. and H. Summala. 1998. Attention and Expectation Problems in Bicycle–car Collisions: An In-depth Study. *Accident Analysis and Prevention, 30*, 657–666.
Ranney, T. 1994. Models of driving behaviour. A review of their evolution. *Accident Analysis and Prevention*, 26(6), 733–750.
Rouse, W. B., & Morris, N. M. 1985. *On looking into the black box: Prospects and limits in the search for mental models* (DTIC #AD-A159080). Atlanta, GA: Center for Man-Machine Systems Research, Georgia Institute of Technology.
SAE. 2014. Taxonomy and Definitions for Terms Related to On-road Motor Vehicle Automated Driving Systems. *SAE Standard J3016*, USA.
Sagberg F., Selpi, G. F. Piccinini and J. Engström. 2015. A Review of Research on Driving Styles and Road Safety. *Human Factors* 57(7):1248-75. doi: 10.1177/0018720815591313
Senders, J. W. 1964. The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. *IEEE Transactions on Human Factors in Electronics*, 5: 2– 6.
Senders, J. W. 1983. *Visual Sampling Processes*. Hillsdale: Erlbaum.
Seppelt, B.D. and J. D. Lee. 2015. Modeling Driver Response to Imperfect Vehicle Automation. *Procedia Manufacturing* 3: 2621-2628. doi. 10.1016/j.promfg.2015.07.605
Seppelt, B.D. and T. W Victor. 2016. Potential Solutions to Human Factors Challenges in Road Vehicle Automation. In Meyer, G. and Beiker, S. (Eds.) *Road Vehicle Automation 3*. Berlin: Springer.
Seth, A.K., 2013. Interoceptive Inference, Emotion, and the Embodied Self. *Trends in Cognitive Sciences* 17: 565–573. doi: org/10.1016/j.tics.2013.09.007
Schultz, W., P. Dayan and P. R. Montague. 1997. A Neural Substrate of Prediction and Reward. *Science* 275: 1593–1599.

Simon, H. A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, *LXIX*: 99–118.
Summala, H. 1988. Risk Control is Not Risk Adjustment: The Zero-risk Theory of Driver Behaviour and its Implications. *Ergonomics*, 31(4): 491–506. doi: org/10.1080/
Summala, H. 2007. Towards Understanding Motivational and Emotional Factors in Driver Behavior: Comfort Through Satisficing. In P. C. Cacciabue (Ed.), *Modeling Driver Behavior in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems*. London: Springer.
Summala, H. and M. Räsänen. 2000. Top-down and botttom-up processes in driver behaviour at roundabouts and crossroads. *Transportation Human Factors*, 2(1): 29 – 37.
Sutton, R.S. and A. G. Barto. 1998. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge.
Theeuwes, J. & Godthelp, H. 1995. Self-explaining roads. *Safety Science*, 19. 217-225.
Vaa, T. 2007. Modeling driver behavior on the basis of emotions and feelings. In P.C. Cacciabue (Ed.), *Modeling Driver Behavior in Automotive Environments: Critical Issues in Driver Interactions with Intelligent Transport Systems*. London: Springer.
Wann, J. P. and M. Land. 2000. Steering with or without the flow: is the retrieval of heading necessary? *Trends in Cognitive Science,* 4: 319-324. doi. 10.1068/v060003.
Wann, J. P., and R. M. Wilkie. 2004. How Do We Control High Speed Steering? In L. M. Vaina, S. A. Beardsley, S. K. Rushton (Eds.), *Optic Flow and Beyond* (371-389). Dordrecht: Kluwer Academic Publishers.
Wolpert, D. M. and R. C Miall. 1996. Forward models for physiological motor control. *Neural Networks* 9(8): 1265–1279. doi: 10.1016/S0893-6080(96)00035-4
Vicente, K. J., & Rasmussen, J. 1992. Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man and Cybernetics*, 22(4), 589-606.
Wickens, C. D. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2): 159-177
Wickens, C. D., J. Goh, J. Helleberg, W. J. Horrey and D. A. Talleur. 2003. Attentional Models of Multi-task Pilot Performance Using Advanced Display Technology. *Human Factors* 45: 360 –380.
Victor, T., Bärgman, J., Boda, C.-N., Dozza, M., Engström, J., Flannagan, C., Lee, J.D.,Markkula, G., 2015. Analysis of Naturalistic Driving Study Data: Safer Glances Driver Inattention, and Crash Risk. SHRP 2 Report S2-S08A-RW-1.Transportation Research Board.
Wilde, G.J.S. 1982. The Theory of Risk Homeostasis: Implications for Traffic Safety and Health. *Risk Analysis* 2: 209-225. doi. 10.1111/j.1539-6924.1982.tb01384