A Hidden Markov Framework to Capture Human-Machine Interaction in Automated Vehicles

C. P. Janssen[a], Linda Ng Boyle[b], Andrew L. Kun[c], Wendy Ju[d] and Lewis L. Chuang[e]

[a]Experimental Psychology, & Helmholtz Institute, Utrecht University, 3584 CS Utrecht, The Netherlands; [b]Industrial & Systems Engineering, University of Washington, Seattle, 98195, USA [c]Electrical and Computer Engineering, University of New Hampshire, Durham, 03824, USA [d]Information Science, Cornell Tech, New York, 10044, USA [e]Max Planck Institute for Biological Cybernetics, Spemannstrasse 41, 72076, Germany

ARTICLE HISTORY

Compiled May 24, 2018

ABSTRACT

We introduce a Hidden Markov Model framework to formalize the beliefs that humans may have about the mode in which a semi-automated vehicle is operating. Previous research has identified various "levels of automation," which serve to clarify the different degrees of a vehicle's automation capabilities and expected operator involvement. However, a vehicle that is designed to perform at a certain level of automation can actually operate across different modes of automation within its designated level, and its operational mode might also change over time. Confusion can arise when the user fails to understand the mode of automation that is in operation at any given time, and this potential for confusion is not captured in models that simply identify levels of automation. In contrast, the Hidden Markov Model framework provides a systematic and formal specification of mode confusion due to incorrect user beliefs. The framework aligns with theory and practice in various interdisciplinary approaches to the field of vehicle automation. Therefore, it contributes to the principled design and evaluation of automated systems and future transportation systems.

CONTACT C. P. Janssen. Email: c.p.janssen@uu.nl

KEYWORDS

Automated Driving; Mode confusion; Handover; Human-Machine Interaction;Semi-autonomous driving; Hidden Markov Models; Automation

1. Introduction

There is rapid progress towards the development of autonomous vehicles (Bengler et al., 2014; Kun, Boll, & Schmidt, 2016). Vehicle automation can reduce the role of the human agent by increasing the responsibilities (also referred to as authority, Flemisch et al., 2012) of the driving task assumed by the vehicle, or non-human agent (Luettel, Himmelsbach, & Wuensche, 2012).

However, autonomous vehicles are unlikely to be effective for all driving situations. Technology is limited in its ability to anticipate all possible traffic situations, particularly for rare events (Gold, Körber, Lechner, & Bengler, 2016). There are also ethical dilemmas (Bonnefon, Shariff, & Rahwan, 2016), and legislation requirements (Federal Automated Vehicles Policy, 2016; Inners & Kun, 2017; Pearl, 2017) that may necessitate a human agent to assume some level of vehicle control, when the non-human agent is unable to perform all aspects of the driving task. This creates a system with shared control between the human and non-human agents.

In systems with shared control, confusion regarding the control authority might arise. Confusion is most prevalent in situations where the control authority shifts due to changes in context. Three aspects of context changes are relevant. First, context changes can happen unexpectedly (e.g., sudden snow, or a "rogue" vehicle). This might leave the human agent with insufficient time to observe and respond appropriately to the context change (e.g., to take over the control of lateral position of the vehicle under snowy conditions).

Second, an automated vehicle's capabilities can change frequently in response to varying context, even within a single trip. For example, its sensing capability can fluctuate due to small pockets of fog, the safety criterion for velocity can change frequently during rush hour, and *ad hoc* modifications to the infrastructure (e.g., construction sites) can change traffic patterns. Such rapid context changes might require the human agent to pay attention and update their knowledge of their immediate surroundings

frequently.

Third, users of autonomous vehicles can be expected to divert their attention away from the driving task as the tasks become more automated (De Winter, Happee, Martens, & Stanton, 2014; Warm, Parasuraman, & Matthews, 2008). This will interfere with their ability to perceive and monitor changes in the vehicle's mode of automation.

All these factors might hinder the human agent in keeping track of context changes, which impact the mode of vehicle operation and the human agent's associated responsibilities. Poor in-vehicle interface design can also increase the likelihood of confusion (Stanton & Marsden, 1996).

Although the research community has recognized the potential for human confusion, the lack of alternatives have fostered research based on a classification scheme for the automation, such as those developed by the Society of Automotive Engineers (SAE International, 2014), the German Federal Highway Research Institute (BASt) (Gasser & Westhoff, 2012), and the US Department of Transportation (National Highway Traffic Safety Administration, 2013) (e.g. Endsley, 2017; Kun et al., 2016; Kyriakidis et al., 2017; Mok, Johns, Miller, & Ju, 2017; van der Heiden, Iqbal, & Janssen, 2017). These frameworks focus on vehicle capability and functions, and capture the operations of an automated vehicle from the system technology perspective, and in identifying what features need to be engaged to accomplish automated driving. In such a view, vehicle control is decomposed into sub-tasks (e.g., lane keeping and cruise control), for which control authority is delegated either to the human or non-human agent, depending on the context (e.g, autopilot might only be enabled on the freeway).

The responsibilities of the human and non-human agent are clear at the extreme ends of such classification schemes. At one extreme is the no automation level where all responsibility is with the human. At the other extreme is the full automation level, where the non-human agent is fully in charge. Yet, a clear-cut division of labor does not exist along the continuum between these extremes for two reasons. First, the classification schemes only identify the number of functions that are automated, not which functions are automated. For example, SAE level-1 automation might refer to a vehicle with cruise control, but it can also refer to adaptive cruise control.

Second, there is the potential of frequent context changes and associated changes

with respect to the responsibilities of the human and non-human agent. Confusion due to context changes has been observed in practice (Endsley, 2017). Moreover, the NHTSA report for the first fatal crash with an automated vehicle, suggests that there was 'a period of extended distraction' (p.9 in Habib, 2017). It is unclear whether the distraction was due to mode confusion. Nonetheless, the long period of distraction contravenes the user requirements of a SAE level-2 vehicle which requires eyes on road. In other words, the user might have acted inappropriately despite the explicit limitations of the system, which suggests a mode confusion.

In order to reduce the odds of mode confusion and, more broadly, to design automated vehicles that allow human and non-human agents to interact safely and effectively, we require a formal framework that articulates how the human agent's beliefs of control responsibilities might vary in response to varying automation across various contexts. Such a framework will serve designers, engineers, and researchers in understanding (through formal explicit specification) differences in the perceived roles of the human and non-human agents across different scenarios. The goal of this paper is therefore to provide a framework that will facilitate the development of systems that will better communicate to the human the mode and limitations of the non-human agent (or vehicle).

2. The Framework: Hidden Markov Models

The starting point of our framework for modeling the relationship between the non-human and human agent are the following explicit assertions: vehicles can operate in different modes of automation (even when designed for a specific automation level) and the human agent's beliefs about the vehicle's modes can be incorrect. For a simple example where adaptive cruise control (ACC) is on or off, Figure 1 illustrates four unique combinations of (non-human) automation ***mode*** and human ***belief***.

Signal Detection Theory (SDT) provides a concise description that explicitly distinguishes two types of incorrect beliefs: a *false alarm* occurs when the human is "over-prepared" and incorrectly believes that action is needed even though automation is *on*, and a *miss* occurs when a human is "under-prepared" and fails to act

due to an incorrect belief that automation is *on* but in fact, the automation is off. More importantly, SDT provides the mathematics for formally deriving the 'sensitivity' of the overall human-automation system, whereby an optimal situation is one that maximizes the number of hits while minimizing false alarms (McNicol, 2005).

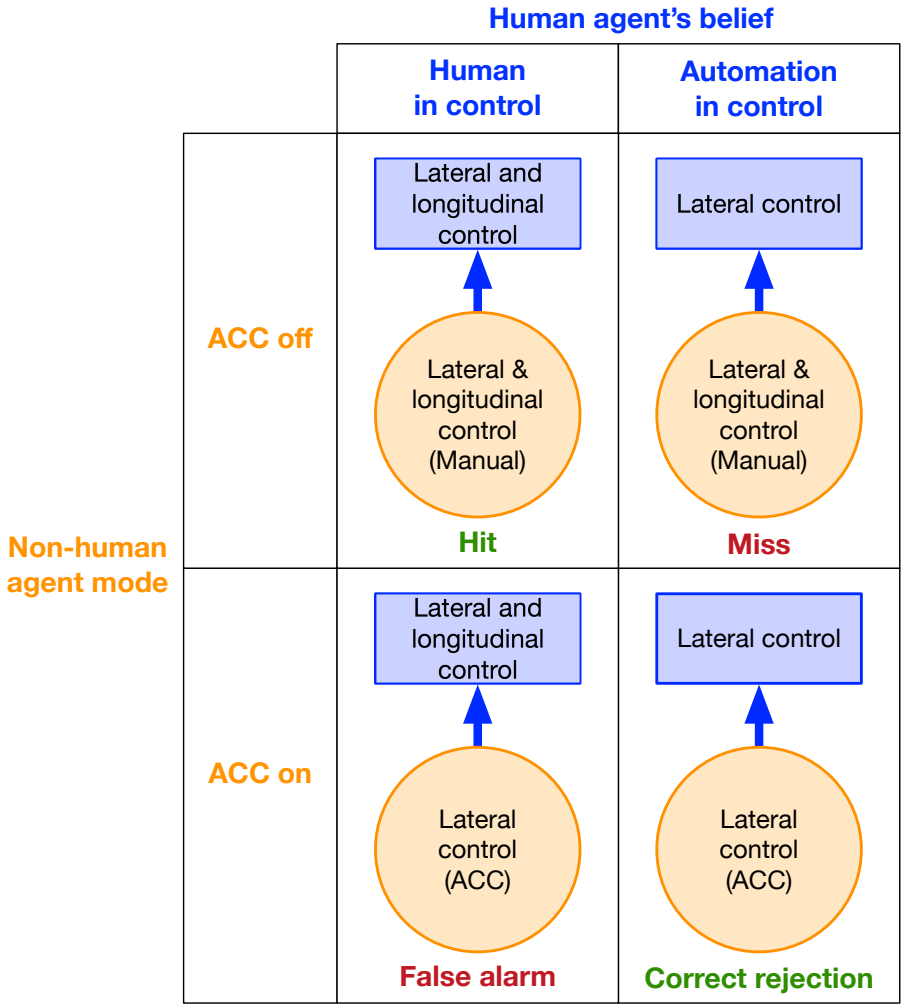

Figure 1. A signal detection characterization of possible human beliefs (squares, varied between columns) of the vehicle's automation mode (circles, varied between rows).

Although SDT allows for a classification of unique combinations of (non-human) agent modes and human agent beliefs and a derivation of sensitivity, it does not account for the variable circumstances that induce each combination. For example, a human may initially (and correctly) believe that ACC is *on* and therefore, they might not

control longitudinal position themselves. However, there are circumstances where the human is in a 'miss' situation, and incorrectly believes longitudinal control is still on when it is not. To capture such dynamics, a framework is needed that can formalize how combinations of human beliefs and automation modes can dynamically change over time and situations, and that can model how likely such changes are.

We propose using ***Hidden Markov Models (HMM)*** (Rabiner, 1989) to capture the dynamic nature of the transitions between the combinations. Similar to SDT, HMMs provide an explicit systems level approach that can formally model the difference between the vehicle's mode of automation and a human user's beliefs over time. In addition, HMMs can account for the uncertainty that is associated with the modes and beliefs.

The framework makes explicit that a vehicle's level of automation is dynamic. In addition, it can represent transitions between different modes in a probabilistic fashion. For example, a vehicle in manual control will remain in manual control with some probability, P_i, and transition to ACC with some other probability, P_j. That is, mode changes can occur over time and context. Furthermore, HMMs assign probabilities to user beliefs associated with each mode of automation. Thus, when the vehicle is under manual control, there is a large probability that the user also believes that it is, in fact, under manual control. Yet, when the vehicle is under ACC control, there is a non-zero probability that the user believes that ACC is on, as well as that the vehicle is under manual control. In traditional HMM terminology, automation modes represent the (hidden or latent) states of a Markov process, while user beliefs represent the probabilistic observations of these states.

The scientific value of an HMM framework is that it allows potential problems and types of confusions to be formally specified. By assigning probabilities to each situation, it allows the likelihood of specific confusions (e.g., of misses and false alarms) to be estimated, thus serving a theoretical or applicational purpose. Reciprocally, empirical data can be used to update these likelihoods, thus providing a common theoretical framework that can benefit from the incremental accumulation of research evidence. Understanding these problems within a systematic framework should also allow system designers to anticipate potential errors and to create principled solutions

and approaches to addressing potential errors that result from user misunderstanding. The value of the HMM framework will now be illustrated with three cases.

3. Cases

3.1. Case 1: Formalizing mode confusion in HMM

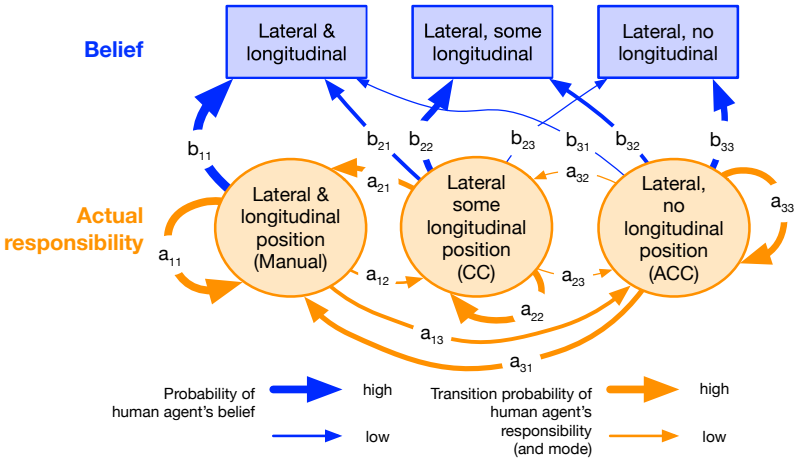

Figure 2. Hidden Markov Model (state transition view) for a car with manual lateral control, conventional cruise control (CC), and adaptive cruise control (ACC). Gold arrows (→) represent the transitions between automation modes and corresponding responsibilities. Blue arrows (→) represent the belief likelihoods that each automation mode might generate in the user. Some system modes can have false beliefs (i.e., mode confusion) associated with them.

This first case (Figure 2) considers a perfect situation in which a vehicle has two optional system functions: conventional cruise control (CC), and adaptive cruise control (ACC) (both functions are classified as SAE level-1 automation; SAE International, 2014). In this HMM, there are three system modes: manual control, CC, and ACC. The system can transition between these modes, with some transitions having higher probability.

The characterization of mode transitions meets the *Markov assumption* that the future mode of the system at time $t + 1$ depends on the current mode at time t. In probabilistic terms, this also means that the probabilities of each transition from each mode together sum up to 1. The transition probabilities from one mode to another in

Figure 2 are marked as a_{ij}, $1 \leq i, j \leq 3$, where the subscripts[1] i, and j indicate the current and next mode respectively. Thus, $\sum_{j=1}^{3} a_{ij} = 1$ for all modes of i .

There are also three beliefs that the human agent can have, corresponding with whether the system is in manual, CC, or ACC mode. However, in line with the signal detection description in Figure 1, beliefs do not need to correspond with the actual mode of the system. This is consistent with empirical observations that human actions (e.g., initiating a mode change) might not always end up in the desired mode, and that the human might not be aware of the resulting mode (Xiong, Boyle, Moeckli, Dow, & Brown, 2012).

Formally, the HMM framework requires that all possible beliefs for a mode sum up to 1. Figure 2 shows the probabilities of beliefs that are possible in any given mode i as b_{ik}, for $1 \leq i, k \leq 3$, for the current mode i, and the belief held by the human agent as k. Thus, $\sum_{k=1}^{3} b_{ik} = 1$ for all modes i.

The contribution of the HMM description is threefold. First, all combinations of modes and beliefs can be expressed in a compact and formal manner. Notice that a SDT description of this framework would require seven unique combinations of beliefs and modes. Second, the framework makes explicit what transitions are possible and how likely these transitions are and, thereby, makes the (hidden) assumptions of the designer or researcher of the system explicit and open to discussion. Third, when probabilities are assigned to the various modes and beliefs, the framework allows one to calculate the likelihood of false alarms and misses.

3.2. Case 2: Formalizing mode confusion in a commercially available system

Our second case describes the user experiences of a commercially available vehicle at SAE automation level 2, the Tesla Model S. A Tesla model S provides adaptive cruise control (ACC) as well as automated lane following (auto steer). This results in four possible user beliefs concerning whether these systems are 'on' or 'off' (see Figure 3).

In an ideal world, there are also four system modes (a, b, c1, d in Figure 3). ACC

[1]The subscripts i and j are bound between 1 and 3 as there are three modes in this example

(automation mode c1) and auto-steer (d), are both initiated via the same lever, which is connected to the steering wheel. Moving the Cruise/Autopilot lever towards the human driver once starts ACC only (mode c1), moving it towards the operator twice in rapid succession leads to ACC with auto steer (mode d). Pressing the brake pedal will always move the system to manual control, while manipulating the steering wheel when ACC and auto steer are engaged will move the system to ACC[2].

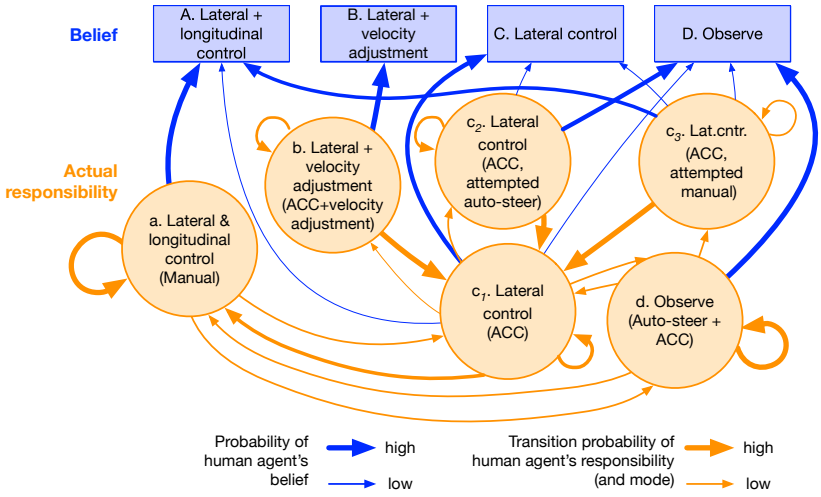

Figure 3. HMM framework (state transition view) that formalizes mode confusion in a level-2 automation system. Note that one type of automation (ACC on) can be associated with multiple modes (c1,c2,c3), and, therefore, with multiple beliefs. Gold arrows (→) represent the transitions between automation modes and corresponding responsibilities. Blue arrows (→) represent the belief likelihoods that each automation mode might generate in the user.

In spite of these distinct interaction designs, mode confusion can occur (e.g., Endsley, 2017). For example, users might attempt to go to manual (mode a) by manipulating the steering wheel, but instead end up in ACC (automation mode c3). In this mode, the user might have the incorrect belief (false alarm in SDT terminology) that they have manual control over all aspects of the vehicle (belief A). Conversely, the user might also attempt to press the lever twice to start ACC and auto-steer (mode d) but not succeed because they only manage to press the lever once. In the resulting mode (c2) they might incorrectly believe that they can simply observe the car (belief D; a miss in SDT terminology).

This case demonstrates how the HMM framework can capture mode confusion in

[2]A user demonstration of this system can be seen here: (Teslavangeliste, 2015).

commercially available systems. The framework makes it very explicit that system modes which are colloquially thought of as one mode, or one level of automation, can in practice be divided into multiple modes. For example, colloquially one might describe a system that incorporates Adaptive Cruise Control and auto-steer as a SAE level-2 vehicle. However, this system has at least six modes, of which five (all but mode d) do not meet the description of a typical SAE level-2 automation (i.e., with the vehicle controlling lateral and longitudinal position on specific road segments).

3.3. Case 3: Including the context of space and time using the lattice view

The previous two cases represent a "state transition" view of the system modes. It is also important to consider the changes in modes and beliefs as the context changes temporally and spatially. A contextual description can be created with the *lattice view* of a Hidden Markov Model, as done in Figure 4. The figure focuses on three modes from Figure 3 and shows how the system mode changes as a result of human action. For example, imagine a driver that starts with ACC on (mode c1), wants to go to Auto-steer mode and presses the Cruise/Autopilot lever, however, only one press is registered. The human agent might believe that their responsibility is to observe (belief D), whereas the car in fact is NOT in auto-steer and lateral control is needed (mode c2). Now, the human belief might update when the human *observes* that the car does not stay in the lane, for example when driving a curvy trajectory. This updates their belief, and might trigger them to subsequently press the Cruise/Autopilot lever twice again, to correctly go to a mode in which they can only observe (mode d, belief D). This case demonstrates how the lattice view keeps track of how the system transitions over time and space.

The lattice view provides insights on the transitions that can be made in the context of all possible transitions during that time period and connects it to the context. Such a contextual description is needed to make predictions for specific roads, traffic, and environmental conditions, and connects the results with the in-situ measurements of human behavior and system functioning. Moreover, a contextual description over time and space aligns with process oriented models of driver behavior and thought from cognitive psychology (e.g. Brumby, Janssen, Kujala, & Salvucci, 2018). Therefore, it

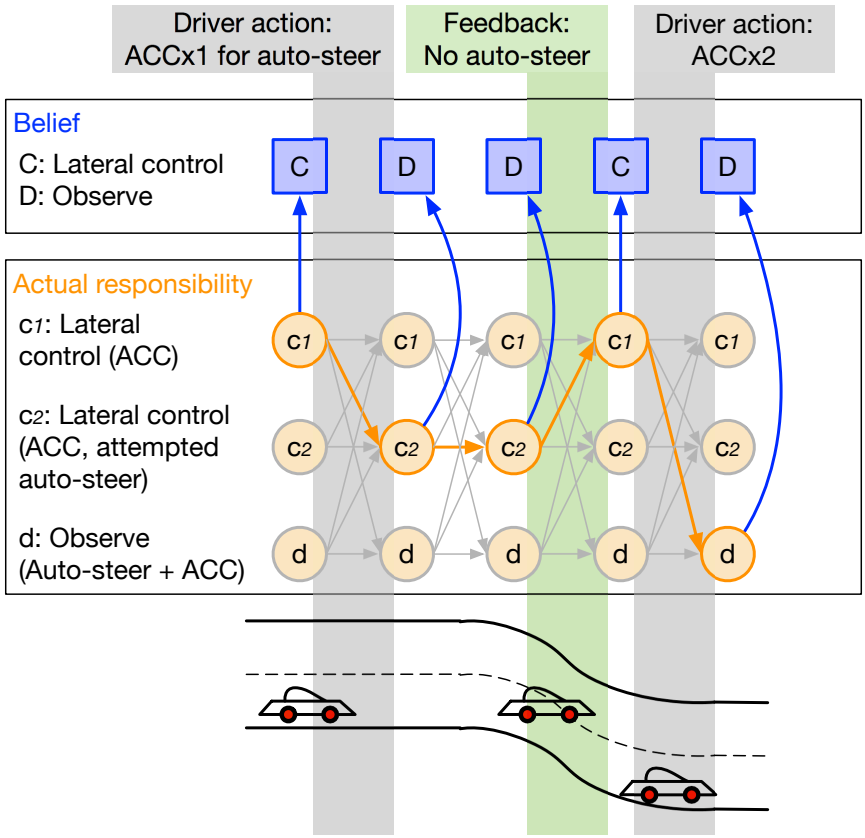

Figure 4. HMM framework (lattice view) of beliefs about lateral and longitudinal control in a specific driving context (intersection, sketched at top). Various paths are possible over time for beliefs and responsibilities (→). In this context a specific path of responsibilities (→) does not always align with beliefs (→) . User action (gray areas) and feedback from the environment (green area) can change the modes and/or beliefs.

more naturally allows for modeling of human behavior over time.

4. General Discussion

We introduced Hidden Markov Models as a framework for representing the distribution of responsibilities between automated vehicles and human drivers in probabilistic terms. The framework provides a novel perspective on mode confusion in four ways. First, the framework makes explicit that vehicle automation *modes* can differ from a user's *beliefs*, as illustrated in the three case studies. Second, the framework makes explicit that modes and beliefs are multidimensional and change over the context (including space and time). Third, the flexibility in representational form makes the framework versatile. Specifically, it can both be used in general system design and analysis using the state transition view, or tied to specific contexts using the lattice

view. Fourth, the consideration of multiple modes and beliefs and their transitions allows for a probabilistic representation of mode confusion. As a consequence, it provides a useful tool to assist in the measurement, estimation, and inference regarding the likelihood of specific errors.

The framework helps to formalize potential problems and types of confusions that can exist. In its application, the HMM framework allows for a more systematic evaluation and proposed mitigation. For example, researchers can use the framework to gather data on how often specific transitions occur and, hence, infer the likelihood of human beliefs (e.g., by observing whether actions are made that are inconsistent with the correct belief). Such empirical work combined with a formal framework can be used to calculate or simulate the likelihood of actions that are inconsistent with the correct belief. The lattice view (case 3) is particularly relevant given the emphasis on context and its natural alignment with psychological process-oriented models of human behavior and thought, including driver distraction (Brumby et al., 2018).

The three example cases illustrated a subset of the human beliefs and system modes, as this was enough to demonstrate the basics of the framework (case 1), how this can be applied to current commercially available systems (case 2), and how this applies to specific contexts (case 3). The cases thereby illustrate the value for an interdisciplinary field. The claim is not that the descriptions completely describe all modes, beliefs, and transitions. This begs the question of when one can be certain that a formal model is complete. This philosophical question can be considered as a strength of the approach: it makes the *assumptions* (of the researcher or designer) behind the system explicit and therefore open for debate.

A pragmatic attempt to have a complete model can be achieved in four steps. First, start with the system design and formalize all known system modes and transitions, including an exploration of possible human actions and whether these change the mode (cf. Thimbleby, 2007). Second, assume that each mode has its own, unique correct belief, yet that all other beliefs are possible alternatives for each mode. Third, anticipate unknowns by also including a belief for "other" to capture all unknowns (i.e., similar to incorporation of an error term in statistical models). Fourth, use a method of choice to prune false beliefs from the system, including consideration of

resilience methods to prevent false beliefs. This fourth step can be done in a variety of ways, including empirical studies, estimates based on other theories or models, design sketches, or expert analysis. The benefit of this fourth explicit step is again that it makes explicit *why* specific beliefs were pruned and that it forms solid documentation for external evaluation of a system design (e.g., a safety assessment).

4.1. Interdisciplinary applications

Applying Hidden Markov Models to capture transitions of control and mode confusion in automated systems is novel, yet builds on a long tradition of Markov Models and probabilistic models, including their use in other human-computer interaction settings (Thimbleby, 2007). Moreover, it naturally fits with theory and practice in multiple disciplines, and thereby provides a useful tool to avoid being lost in translation and to facilitate convergence in the field. Below we make suggestions of use for specific fields.

For engineering, computer science, and system design, using HMMs aligns with a systems approach to design, in which system states (including automation modes) and their transitions are explicitly identified. Existing frameworks of automation (e.g. Gasser & Westhoff, 2012; National Highway Traffic Safety Administration, 2013; SAE International, 2014) do explicitly distinguish between the various stages of automation. However, in contrast to the HMM framework, they do not currently incorporate human beliefs and express them explicitly.

For psychology, human factors, and human-computer interaction, the HMM framework allows a formal way to express theories of human behavior. Theories that naturally fit with the framework are hierarchical decomposition of tasks (Card, Moran, & Newell, 1983), including decomposition of human driving (Michon, 1985), and probabilistic notions of human beliefs and Bayesian belief updating given observations of actions (Baker, Jara-Ettinger, Saxe, & Tenenbaum, 2017). Although such theories are already combined in process-oriented models of driver behavior and thought (Brumby et al., 2018), the HMM framework is explicitly framed in a way that aligns more naturally with fields outside of psychology and, in principle, can be used both for in-situ measurement (using the lattice view) and systems design. This provides room for more cross-fertilization.

For the field of design, the framework can make explicit the confusion that can arise for which 'automation modes' and in which 'context mode'. Subsequently, this focuses research to address the specific interventions that have to be adopted in order to avoid these specified confusions. Specifically: if the framework identifies a potential miss or false alarm, how can design be used to mitigate that error?

Finally, for the assessment of driving safety, including by governments, the HMMs provide a tool for assessing the safety of systems in general (using the state transition view) or in specific contexts (using the lattice view). For example, such an assessment can focus on the complexity of the system (e.g., how many modes are there? what transitions are possible?), as well as on the anticipation of confusion in human beliefs (e.g., what misses and false alarms have been identified and mitigated? how well have those been mitigated? which misses and false alarms have been left out compared to a full system that associates all beliefs with all modes?).

4.2. Extensions and future work

The potential for interdisciplinary use of the HMM framework also highlights several opportunities for further work. We highlight six paths. First, the complexity of a Hidden Markov description of mode confusion needs to be further investigated. In general, the maximum complexity of a system with N modes is $2 \times (N+1)^2$. The assumption behind this equation is that there are N modes between which there can be transitions, including self-transitions. As there might also be unknown modes, we allow one mode unknown to capture all these other modes, resulting in $(N+1)$ modes. The maximum number of transitions between these modes (including self-transition) then becomes $(N+1)^2$. Similarly, in the maximum complexity case, each mode of the world (including unknown) has its own belief associated with it. Also, in the maximum complexity case, each possible mode might be connected with each possible belief, again giving $(N+1)^2$ edges. The maximum complexity for any given number of modes N, as estimated by the number of edges, then becomes $2 \times (N+1)^2$. As we described earlier, in a pragmatic account there are multiple ways to prune this tree. For example, in our case 1 (Figure 2), we pruned the maximum space of $2 \times (3+1)^2 = 32$ edges to only 16 edges by not including an "other" mode and by ruling out unlikely

mode-belief connections, such as the belief that the car is in ACC mode when it is in manual (and would slow down if the driver did not press the gas pedal).

Second, empirical evidence can be gathered to estimate the exact probabilities of making a false belief. The HMM framework clarifies the modes, beliefs, and transitions that ought to be considered, thus allowing for empirical investigations to be consistent in their endeavors. The identification of these probabilities can be followed by estimation and calculation. Even though there is no empirical estimate of exact probabilities, there is empirical evidence of mode confusion in semi-automated vehicles (e.g. Endsley, 2017; Mok, Johns, Gowda, Sibi, & Ju, 2016).

Third, the framework would be useful in the application of sensor technology to assess user state (e.g., eye-tracking for situational awareness) for varying system modes and human beliefs. A system might be designed to assess the most likely belief of the user in a probabilistic manner and to act on this uncertainty (e.g., for example using a Partially Observable Markov Decision Process, see Howes, Chen, Acharya, & Lewis, 2018). For example, if the system detects that the user does not respond to lateral sway (i.e., green segment in Figure 4), the car might infer that the user might not have the appropriate belief about the system mode. The actions that are appropriate for the non-human system is an important research question, but can include explicit communication of system mode (e.g., an alert), acting on the appropriate mode (i.e., making a steering correction), and/or changing the system mode to better facilitate the user's beliefs.

Fourth, the framework can be used to further study user beliefs in relation to trust. Dynamically adjusting the system mode, given an understanding of the changes in user's beliefs over time, might also be crucial given their trust in the automation (Abe & Richardson, 2006; Bliss & Acton, 2003; Riener, Boll, & Kun, 2016; Wickens & Dixon, 2007). Like user beliefs in general, trust in automation is no longer a chronic user state that is dependent on a fixed level of system reliability. Instead, it could now be considered in terms of expected user engagement, given the conditional likelihood of mode transitions in the automation. An automated system with a high likelihood of mode transitions need not be considered to be unreliable. However, it would mandate more user engagement, the lack of which would be reflected as mode confusion.

Fifth, it is an open question what platforms are best for the design of HMMs regarding human-automation interaction. Trade-offs for such efforts are described more widely (including example code) in the book by Thimbleby (Thimbleby, 2007) for the general domain of human-computer interaction.

Sixth, and finally,the application to high levels of automation can be considered. For our framework, the potential confusion between user beliefs and system modes is crucial. Even though some systems are designed to remove the human completely, such as the Waymo (Google) autonomous car (i.e., they expect to achieve SAE level 5 automation), the human operator will still need to provide some input and updating. For example, it might be necessary to inform the user about the current trajectory of the drive (e.g. should the car take the "fast" or "scenic" route?). Although this may appear as if human safety is not impacted, possible mode confusion needs to be minimized regardless to enhance overall user experience (Kun et al., 2016).

5. Conclusions

In conclusion, this paper introduces Hidden Markov Models as a formal probabilistic framework for describing the uncertainty about the combination of vehicle mode and human beliefs, and the transitions between them. It is applicable to any situation where some form of shared responsibilities between a human and a computer system exist, even beyond the automotive context. The value of the work lies in the formal representation of modes and beliefs, which uncovers hidden assumptions and hidden problems in a way that can be alleviated by multiple disciplines.

Acknowledgements

This paper idea was generated at the seminar on Automotive User Interfaces in the Age of Automation held at Schloss Dagstuhl, Germany (Seminar number 16262, Riener et al., 2016). We gratefully acknowledge the organizers of the seminar, Schloss Daghstul, and the other members of the seminar for their contributions.

Funding

Christian P. Janssen was supported by a Marie Sklodowska-Curie fellowship of the European Commission (H2020-MSCA-IF-2015, grant agreement no. 705010, 'Detect and React'). Lewis Chuang was supported by the German Research Foundation (DFG) within the project C03 of SFB/Transregio 161.

References

Abe, G., & Richardson, J. (2006). Alarm timing, trust and driver expectation for forward collision warning systems. *Applied Ergonomics*, *37*(5), 577–586.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*, 0064.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 6–22.

Bliss, J. P., & Acton, S. a. (2003, nov). Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied Ergonomics*, *34*(6), 499–509.

Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016, June). The social dilemma of autonomous vehicles. *Science*, *352*(6293), 1573–1576.

Brumby, D. P., Janssen, C. P., Kujala, T., & Salvucci, D. D. (2018). Computational models of user multitasking. In A. Oulasvirta, P. Kristensson, X. Bi, & A. Howes (Eds.), *Computational interaction design.* Oxford: Oxford University Press.

Card, S. K., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

De Winter, J. C., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, *27*, 196–217.

Endsley, M. R. (2017). Autonomous driving systems: A preliminary naturalistic study of the Tesla Model S. *Journal of Cognitive Engineering and Decision Making*, *11*(3), 225-238.

Federal Automated Vehicles Policy. (2016). *Accelerating the next revolution in roadway safety*

(Tech. Rep.). National Highway Traffic Safety Administration, Department of Transportation.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2012). Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work*, *14*(1), 3–18.

Gasser, T., & Westhoff, D. (2012). *BASt-study: Definitions of Automation and Legal Issues in Germany* (Tech. Rep.). Federal Highway Research Institute (BASt). Retrieved from `http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/Gasser.pdf`

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations. *Human Factors*, *58*(4), 642-652.

Habib, K. (2017). *Automatic vehicle control systems* (Tech. Rep. No. PE 16-007). National Highway Traffic Safety Administration, Department of Transportation.

Howes, A., Chen, X., Acharya, A., & Lewis, R. L. (2018). Interaction as an emergent property of a partially observable markov decision process. In A. Oulasvirta, P. Kristensson, X. Bi, & A. Howes (Eds.), *Computational interaction design.* Oxford: Oxford University Press.

Inners, M., & Kun, A. L. (2017). Beyond liability: Legal issues of human-machine interaction for automated vehicles. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 245–253). New York, NY, USA: ACM.

Kun, A. L., Boll, S., & Schmidt, A. (2016). Shifting gears: User interfaces in the age of autonomous driving. *IEEE Pervasive Computing*, *15*(1), 32–38.

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., ... Happee, R. (2017). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 1–27.

Luettel, T., Himmelsbach, M., & Wuensche, H.-J. (2012). Autonomous ground vehiclesconcepts and a path to the future. *Proceedings of the IEEE*, *100*(Special Centennial Issue), 1831–1839.

McNicol, D. (2005). *A primer of signal detection theory.* New York, NY: Routledge.

Michon, J. A. (1985). A critical view of driver behavior models: What do we know, what should we do? In L. Evans & S. R. C (Eds.), *Human behavior and traffic safety* (pp. 485–520). Boston, MA: Springer.

Mok, B., Johns, M., Gowda, N., Sibi, S., & Ju, W. (2016). Take the wheel: Effects of available modalities on driver intervention. In *Intelligent vehicles symposium, 2016 ieee* (pp. 1358–1365). Gothenburg, Sweden: IEEE.

Mok, B., Johns, M., Miller, D., & Ju, W. (2017). Tunneled in: Drivers with active secondary tasks need more time to transition from automation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2840–2844). New York, NY, USA: ACM.

National Highway Traffic Safety Administration. (2013). *Preliminary statement of policy concerning automated vehicles* (Tech. Rep.). Washington, DC.

Pearl, T. H. (2017). Hands on the Wheel: A Call for Greater Regulation of Semi-Autonomous Cars. *Indiana Law Journal*(March), 1–47.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Riener, A., Boll, S., & Kun, A. L. (2016). Automotive User Interfaces in the Age of Automation (Dagstuhl Seminar 16262). *Dagstuhl Reports*, *6*(6), 111–159. Retrieved from `http://drops.dagstuhl.de/opus/volltexte/2016/6758`

SAE International. (2014). *J3016: Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.*

Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. *Safety Science*, *24*(1), 35–49.

Teslavangeliste. (2015). *Tesla model s adaptive cruise control explanation and demonstration 70d: https://www.youtube.com/watch?v=yzo5pjelmne.* Retrieved from `https://www.youtube.com/watch?v=yZO5PjeLmnE`

Thimbleby, H. (2007). *Press on: principles of interaction programming.* Cambridge, MA: MIT Press.

van der Heiden, R. M., Iqbal, S. T., & Janssen, C. P. (2017). Priming drivers before handover in semi-autonomous cars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 392–404). New York, NY: ACM.

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors*, *50*(3), 433–441.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: a synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201–212.

Xiong, H., Boyle, L. N., Moeckli, J., Dow, B. R., & Brown, T. L. (2012). Use patterns among

early adopters of adaptive cruise control. *Human Factors*, *54*(5), 722–733.

Biographies

Christian P. Janssen is an assistant professor of experimental psychology at Utrecht University. He received his PhD in human-computer interaction from UCL (2012). His background is in cognitive science, HCI, and artificial intelligence. His major research interests are in multitasking and (driver) distraction, including in automated settings.

Linda Ng Boyle is a professor in industrial & systems engineering at the University of Washington. She received her PhD in civil & environmental engineering at UW (1998).

Andrew L. Kun is an associate professor of Electrical and Computer Engineering at the University of New Hampshire. He received his PhD in Electrical Engineering from the University of New Hampshire (1997).

Wendy Ju is an assistant professor of information science at Cornell Tech. She works on designing interactions with automation, with a focus on human-robot interaction and autonomous vehicle interaction. She holds a PhD in Mechanical Engineering from Stanford, and a MS in Media Arts & Sciences from MIT (2008).

Lewis L. Chuang is a research group leader at the Max Planck Institute for Biological Cybernetics. He received his PhD in Behavior Neuroscience from the Eberhard-Karl University of Tübingen (2011). His research interests are in human attention and information processing during closed-loop interactions with machines.