Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-022-01286-2 OPEN

How should autonomous vehicles drive? Policy, methodological, and social considerations for designing a driver

Amitai Y. Bin-Nun [1]✉, Patricia Derler[2], Noushin Mehdipour[1] & Radboud Duintjer Tebbens[1]

Autonomous vehicles (AVs) are being developed, tested, and commercialized around the world. While skilled human drivers can rely on their experience and common sense to navigate complex driving situations that involve trade-offs between competing objectives, AVs are engineered systems, which may handle complex scenarios based on driving principles articulated at the time of system design. This raises the question of what constitutes proper driving behavior in a complex driving scenario. Many jurisdictions point to existing rules of the road as a description of good driving and, by requiring AVs to follow such rules, hope to improve the safety and efficiency of the transportation system. This paper discusses the desirability of a comprehensive definition of AV behavior, reviews subnational, national, and international regulatory developments that seek to define how AVs might drive, and discusses the tensions between safe, lawful, and efficient driving. Locally defined rules of the road can serve as a guide to a comprehensive driving behavior specification. However, translating rules of the road, which are legal documents written in natural language, to formal rules for use by computers deployed on AVs is a challenging task. In particular, the pervasive appeals to judgment that are present in many rules of the road do not easily lend themselves to the precise formalization of conditions and quantification of values that computers use to make decisions. This work also considers the effect that formalizing behavior for adoption by AVs might have on the general driving culture, and especially on the relationship between existing classes of road users. To highlight the challenges associated with formalizing the rules of the road, this work reports on an experiment where two teams independently translated two rules of the road into formal rules to instruct AVs or to verify the correctness of AV behavior. The study results emphasize the desirability of new technical and political structures to mediate a shared understanding of the rules of the road. The harmonization of behavioral expectations has the potential to improve the safety and efficiency of AV deployments, as well as the broader transportation system.

[1] Motional, Boston, MA, USA. [2] Kontrol, AT and Palo Alto Research Center (PARC), Palo Alto, CA, USA. ✉email: amitai.binnun@motional.com

Introduction

Autonomous vehicles (AVs) have the potential to bring significant benefits to society, including improved safety, accessibility, energy efficiency, land use, and affordability of transportation systems (Bin-Nun and Binamira, 2020; Claypool et al., 2017; Feen et al., 2020; Kalra and Groves, 2017; Taiebat et al., 2018b). Among the challenges that remain in deploying AVs at scale is ensuring that their performance meets societal expectations for safety, lawfulness, and utility. A considerable amount of work is ongoing, including collaborative industry approaches, to create standardized approaches to building and deploying safe AVs (Automated Vehicle Safety Consortium, 2022; ISO, 2018, 2020; Koopman et al., 2019; Wood et al., 2019).

This article focuses on some of the unique challenges in designing AV driving behavior. These challenges emerge because AVs fuse the traditional role of the vehicle manufacturer with what has traditionally been the role of the human driver. AV developers design and construct not only the physical vehicle and systems for executing requested driving behavior (e.g., steering, braking), but also design and build systems that make driving decisions.

There are mature engineering practices for designing a hardware and software system that correctly and reliably executes a well-defined task. In the automotive industry, there are well-developed standards used for this purpose (ISO, 2018). Governments, usually through national bodies such as the United States National Highway Traffic Safety Administration (NHTSA), or international bodies such as the United Nations Economic Commission for Europe (UNECE) or the European Union, expect original equipment manufacturers to develop systems in accordance with highly specific regulatory requirements.

Governments have generally recognized a compelling interest in harmonizing regulatory requirements for vehicle systems as broadly as possible; if each city adopted their own regulatory requirements (e.g., one city required rear-seat airbags and another did not permit them), then the same vehicle might be legally sold and used in one city, but not in a neighboring one. Reflecting a societal interest in avoiding a highly fragmented market for vehicles, the United States has put in place legislative and judicial measures that can preempt subnational regulation of vehicle systems (Haas, 2001). A desire to coordinate regulatory vehicle system requirements across countries motivates the UNECE's World Forum for Harmonization of Vehicle Regulations (Working Party 29). Working Party 29 implements vehicle regulations under the authority of three international agreements, with dozens of countries signing as contracting parties for some or all of these agreements, representing most of the world's population (Chakraborty et al., 2020).

In contrast to regulatory requirements for vehicle systems, legal requirements governing driving behavior traditionally apply to individual drivers, not vehicle manufacturers. A NHTSA study (Singh, 2015) showed that errors in human performance (including errors in scene recognition, decision-making, or execution) contribute to the overwhelming majority of crashes. Thus, an AV that is capable of perfectly executing a planned driving behavior might still be unsafe if the executed behavior is not safe. Therefore, the design of AV driving behavior is key to AV safety (De Freitas et al., 2021). However, formulating safe driving behavior for an AV presents substantial and intertwined engineering and policy challenges. There are significant difficulties in treating driving behavior using traditional engineering methods, in part due to the gap between how policy treats vehicle system design and driving behavior. Particularly in the United States, the local authorities who own the roads generally also determine the rules of the road (ROTRs), with ROTRs varying by state, county, and municipality (Smith, 2013, 2017). If AV driving behavior is part of its system design, then this raises the question of whether that behavior should be regulated by local authorities (as is the case in the United States), or by the policymakers—at the national or international level—responsible for setting regulatory requirements for vehicle system design.

Additionally, the very act of formalizing behavior—by developers, governments, and civil society—will have a significant impact on the interaction and hierarchy of different road users. A juxtaposition of laws, advisory documents, and cultural norms contribute to the interaction between different road users; the act of designing and fixing behavior has the potential to "reorder the culture and concrete of our roads, by flattening the multi-dimensional rules of the road, hardening rules that are currently soft and standardizing across diverse contexts" (Tennant et al., 2021). The current balance of power between different users on the road is deeply shaped by societal discourse between various interest groups (Schmitt, 2020), with historically an important role played by automotive interests in setting expectations for non-automotive road users (Norton, 2011).

The behavior of AVs could reinforce or reset the relationship between road users, with potentially profound implications for non-automotive road users in urban contexts (Latham and Nattrass, 2019). Some have argued that the implications of AV behavior for multiple stakeholders necessitates that the definition of good driving for AVs be determined through democratic institutions and elected representatives (Reed et al., 2021).

This paper discusses the role of behavioral specification, or the formal description of desired AV driving behavior, in the context of current policies and engineering practices. Our primary frame is to understand the implications of current laws for setting AV behavior, identifying gaps, and contributing towards a methodology for stakeholders to better collaborate on specifying AV behavior. First, we review the importance of formulating good driving behavior for AVs and current policy considerations for formulating AV behavior, including nascent efforts by global policymakers to update current vehicle regulatory frameworks to include driving behavior specifications. We then discuss the feasibility of using ROTRs as a foundational set for deriving good driving behavior, current industry efforts to formalize ROTRs for implementation in an AV or evaluation of an AV's driving behavior, and current methodologies for deriving AV behavioral specifications from ROTRs and other sources. We then consider how choices in behavior formalization—including the very act of formalization itself—has potential to impact the nature of the road as public space and the hierarchy of relationships between road users. We report the results of an experiment where two groups worked in parallel to independently translate two ROTRs from the State of Nevada, United States, into formal rules. The experiment demonstrated that ROTRs, as currently formulated, do not easily lend themselves a comprehensive behavioral specification for AVs. The final section proposes steps to address current gaps and identifies technical and policy steps to make progress in resolving the identified challenges and promote safer AV driving behavior.

Using rules of the road to specify autonomous vehicle driving behavior

This section discusses the role that legal ROTRs currently play in policies and standards for AVs.

Why rules of the road? AVs are being developed for multiple use cases. Particularly for urban applications such as autonomous on-demand mobility services, AVs encounter complex driving scenarios with multiple road actors. To help ensure that AVs are

designed correctly, industry participants continue to use and further develop automotive system engineering frameworks (ISO, 2018, 2019, 2020). Developers face the challenge of designing the AV system's driving behavior to meet stakeholder expectations. To build a system that drives well and according to expectations, developers need to know what those expectations are with as much detail as possible. This brings into sharp relief the policy side of this question: Who determines what correct driving behavior is and whose values carry the day? Who are the stakeholders with a say in the matter and what is the process for determining what it means for an AV to be a good driver? And, as we will explore in a later section, what are the implications of introducing designed behavior into the traffic ecosystem and its existing hierarchies?

A natural candidate for defining good driving behavior is ROTRs. ROTRs are specifically designed to prevent the conditions that lead to crashes (Blais and Dupont, 2005), with a study finding that driver violation of laws and norms is more strongly linked to crashes than driver errors (Parker et al., 1995). In many jurisdictions, such as in the United States, policymakers create ROTRs in response to both local conditions and constituent interests. While human decisions to violate ROTRs seem to be driven in part by a cost-benefit evaluation that includes the likelihood of being caught and penalized, AVs offer the opportunity to improve road safety by reducing the potential for drivers to violate ROTRs for impulsive or self-interested reasons (Yagil, 2005).

Autonomous vehicles and rules of the road in policy. As AVs enter service in greater numbers, government bodies at the international, national, and subnational levels continue to create policies aimed at facilitating their safe integration into the road transportation system. Policies focus on a broad set of areas, from permitting and insurance requirements to minimum safety and consumer protection requirements for operating an autonomous ride service (Brown et al., 2018; Channon et al., 2019).

In this context, several entities have suggested that AVs adapt ROTRs to govern their behavior. In the United States, while federal policy does not typically engage ROTRs, NHTSA has "encouraged [developers] to have a documented process for the assessment [of AVs]... obeying traffic laws [and] following reasonable road etiquette", but clearly delineates ROTRs as a state responsibility (NHTSA, 2017). Certain states, such as Nevada, require AVs to comply with ROTRs (Nevada Legislature, 2022a). In 2019, the Uniform Law Commission, a body that seeks to harmonize state laws within the United States by drafting model legislation, finalized its model legislation for AVs (National Conference of Commissioners on Uniform State Laws, 2019). The model legislation recommends that AV developers provide "sufficient evidence" that the AV is "capable of complying with traffic laws," which, according to the Uniform Law Commission, provides "flexibility" for developers acting in "good faith". It also recommends that state ROTRs "be interpreted to accommodate the development and deployment of automated vehicles in a way that maintains or improves traffic safety". These statements can be seen as an implicit recognition of the challenges that would be posed by requiring AVs to strictly and inflexibly comply with ROTRs. Additionally, policy flexibility on AV compliance with ROTRs dovetails with the perspective that an inflexible requirement to comply with ROTRs may not best serve the ultimate goal of promoting the safe integration of AVs into the road transportation system (Smith, 2017).

The Alliance for Automotive Innovation, a prominent United States-based trade association of automotive manufacturers, called for alignment of ROTRs between states and synchronization with international standard bodies (Alliance for Automotive Innovation, 2022). Ontario's pilot deployment requires that AVs follow "all current Highway Traffic Act rules of the road" (Ministry of Transportation, 2022). In Germany, a recently adopted law (Federal Ministry of Transport and Digital Infrastructure, 2021) allows for the deployment of AVs. The law recognizes that human drivers occasionally need to violate certain ROTRs and that implementing the judgment to do so in an AV may be difficult. In Austria, a 2019 amendment to a 2016 framework for the testing and use of AVs explicitly requires compliance with the Austrian road traffic act as well as other relevant laws (Federal Minister for Transport, Innovation and Technology, 2019).

In the United Kingdom, the AV developer Five AI has suggested the creation of a "Digital Highway Code", i.e., an implementation of ROTRs and driving best practices specifically formulated for AVs (Five AI, 2019). The Law Commission of England and Wales and the Scottish Law Commission responded that such a code "may be desirable" but would be "extremely difficult to produce" and alludes to cultural gaps between policymakers and engineers (Scottish Law Commission, 2018). The Commissions recently recommended creating a forum for developers and governments to jointly discuss principles for adapting ROTRs for AVs (Scottish Law Commission, 2020).

Singapore's Land Transportation Authority (LTA) issued Technical Reference (TR) 68, a multi-part, broad, and detailed regulation first published in 2019 (Singapore Standards Council, 2019). TR 68 spells out that some ROTRs do not easily translate for use by AVs and offers a framework for determining when they do or do not (for example, the standard states that the ROTR to check in the rear-view mirror before a lane change does not apply to AVs). TR 68 also contains a section that makes certain ROTRs more formal and implementable for AVs.

The International Telecommunications Union, a United Nations specialized agency for information and communication technologies, led the Focus Group on AI for Autonomous and Assisted Driving (FG-AI4AD). The group identified issues related to post-crash incident handling and information exchange, which are usually governed by a blend of legal and cultural norms. The focus group dubbed this the "Molly problem" and suggested that AV developers may need to address at least some legal considerations that extend beyond the core driving task (Vellinga, 2021).

Behavioral specifications vs. rules of the road. Several policy-making bodies, such as the UNECE (United Nations Economic and Social Council, 2020) and Singapore's LTA (Singapore Standards Council, 2019), have begun the process of writing behavioral specifications for AVs. Behavioral specifications, as defined earlier, are a precise, usually mathematical, embodiment of the driving behavior that the AV is expected to implement. In a sense, behavioral specifications are like ROTRs because they govern on-road behavior. However, behavioral specifications are different in a critical manner—they apply to the developer who builds the AV to execute the specified behavior rather than to a human operator of the vehicle.

The UNECE regulates vehicles for participating countries through Working Party 29. Working Party 29 has a working group for connected vehicles and AVs and has released several behavioral specifications for Level 2 and Level 3 systems (defined in the SAE International levels of driving automation (SAE International, 2018) as vehicles requiring a human driver in the vehicle to either supervise the driving task or serve as a fallback if needed). The UNECE has already finalized regulations that govern the distance at which a Level 2 or Level 3 vehicle should follow another car, the minimum clearance necessary for the vehicle to execute a lane change requested by the human driver,

and the maximum lateral acceleration permitted by automated lane keeping systems (UNECE, 2018; United Nations Economic and Social Council, 2020). These behavioral specifications have the force of law in participating countries and are some of the most precisely formulated behavioral specifications used in a regulatory context.

The work by the UNECE and LTA highlights the distinction between ROTRs and behavioral specifications. ROTRs are written in natural language as they are to be interpreted by human drivers, often using judgment. Behavioral specifications are written in formal mathematical or logical form designed for integration into an engineered product.

Behavioral specifications begin to bridge the gap between ROTRs and traditional systems engineering. Traditional system engineering uses well-defined standards for deriving design requirements for traditional, human-driven vehicles and their subsystems. These requirements include specified performance on attributes such as durability, crashworthiness, security, functionality, failure rates, and other properties; these standards are deeply ensconced within legal, regulatory, and liability frameworks that have well-understood interactions with standard system design methodologies.

In contrast, while driving behavior is also largely governed by legal and regulatory codes, those sources (1) are highly decoupled from the legal frameworks that govern vehicle system design, (2) define correct behavior in a far less objective and reproducible manner than typical system requirements, and (3) are not generated by a methodology that systematically emerges from the desired safety outcome, so it seems unlikely that ROTRs alone can be an exhaustive description of the behaviors necessary for safe driving (De Freitas et al., 2021; Prakken, 2017; Rothengatter, 1997).

Today, behavioral specifications developed by policymakers are fairly limited in scope and do not supersede local ROTRs, so they should be seen as immature and far from comprehensive in their specification of driving behavior. In fact, the very existence of the efforts to formulate behavioral specifications highlights the reality that ROTRs, as currently written in most jurisdictions, may prove too unspecific for straightforward integration into AVs.

Industry standards. The early move towards more rigorous behavioral specifications by policymakers is occurring in parallel with growing industry efforts to develop standards for AVs. Emerging industry best practices and standards also recognize the need for AVs to comply with ROTRs to the greatest practical extent. A white paper from a consortium of AV developers (Wood et al., 2019), which evolved into an International Standards Organization (ISO) Technical Report (ISO, 2020), stated that "machine-interpretable traffic rules are also necessary, as the automated vehicle should obey traffic rules [...] to produce a lawful driving plan, unless exceptions are necessary to prevent collisions". The report specifically calls out creating a "collision-free and lawful driving plan" as a key functionality of AVs and discusses formal rules to encode "explicit traffic rules", "implicit traffic rules", and potentially "hierarchical sets of rules" as a "promising solution" to "challenge[s] in automated driving", particularly the need to "drive in a collision-free manner without compromising comfort or traffic flow" (Wood et al., 2019).

While not specifically focused on formalizing ROTRs, other efforts increasingly focus on the broader topic of AV behavioral specification. Specifically, the Institute of Electrical and Electronics Engineers (IEEE) has released a standard outlining what might constitute reasonable and foreseeable behavior of road users (other than the AV) in certain, specific scenarios; this information could serve as an input for expected AV behavior (IEEE P2846, 2022).

Applications and importance of defining driving behavior. AVs are complex systems consisting of multiple subsystems that contribute to their overall behavior. The need to execute specific behaviors will influence the design and construction of the AV's subsystems. Regardless of the specific implementation of the AV technology, developers may benefit from a way to verify that the AV conforms with the desired behavior and that the subsystems correctly support the behavior. All of these activities may prove easier to accomplish with specific, mathematical descriptions of the desired driving behavior. For instance, creating a collision-free and lawful driving plan can depend on information from perception, prediction, and localization subsystems. An autonomous driving technology consortium report (Wood et al., 2019) discusses this complexity by outlining the various subsystems and their interconnection. Certain information may be necessary to understand if driving is lawful (e.g., location of stop signs, traffic lights, other vehicles, etc.); understanding and enumerating the inputs to determine lawful driving will be helpful to design relevant subsystems so that they can provide the necessary inputs. If ROTRs are to guide AV behavior, casting them in a highly specific, mathematical form would help support these system analyses. Table 1 gives illustrative examples of subsystem and system analysis activities that depend on representation of ROTRs as formal rules. Behavioral specification impacts the development of the entire system and can play a role in a broad range of subsystems and life cycle activities ranging from the development of system requirements to real-time path planning.

Having established the centrality of specifying the desired driving behavior for developing a safe AV from both engineering and policy perspectives, we now turn our attention to ongoing efforts to turn ROTRs into formal rules to guide AV behavior and development.

Formalizing rules of the road

As discussed earlier in this article, a broad range of stakeholders have suggested that AVs comply with ROTRs. This statement carries an implication that it is feasible to determine whether a particular sample of driving does or does not comply with a given ROTR. However, translating natural language into formal, machine-interpretable rules is a complex undertaking. Even sophisticated machine learning and natural language processing methodologies cannot completely automate the translation process (Brunello et al., 2019; Kate et al., 2005). Translating text into formal, mathematical statements ideally captures both the intention of the text and its literal meaning. As discussed, ROTRs as currently written for human drivers often lack the specificity needed for unambiguous evaluation of compliance. To fill this gap, several AV developers have proposed rule-based approaches that include ROTRs as formal rules in AV behavior specifications.

Rules of the road as formal rules. A 2017 study (Prakken, 2017) laid out the importance of both mathematically specifying ROTRs and of embedding these rules into a broader reasoning framework. The study suggested that the absence of such a framework represents a significant gap in AV development.

> *[...] the behavior of autonomous systems should not be seen as rule-governed but as rule-guided. Legal rules are just one factor influencing socially optimal or permissible behavior. Other factors are, e.g., social conventions, individual or social goals or simply common sense. And sometimes these other factors override the legal factors. Having said so, even*

Table 1 Applications and use cases for formal rules.

Use case	Role of ROTR specification	Example
Developing system requirements	Driving involves certain actions by the AV as a response to various properties of other road actors, the environment (e.g., lane lines, crosswalks), and the AV itself. To determine system capabilities and make appropriate design decisions early in the development process, AV developers will likely need to understand what information is needed to comply with formal rules.	A formal rule that checks whether an AV correctly yields to pedestrians at crosswalks informs system capabilities to identify pedestrians and crosswalks.
Planning	A path planning algorithm of an AV can use formal rules to compute trajectories that minimize violation of the rules.	A path planning algorithm computes a trajectory that decelerates smoothly to satisfy a formal rule to stop at a stop sign.
Online verification	Once a path planner has computed a trajectory, an additional subsystem (verification engine) can check whether the trajectory violates any formal rules.	A monitor continuously checks whether the planned trajectory satisfies a formal rule to stay on the road surface and issues a warning in case of potential imminent violations.
Offline verification and validation	AV developers can use formal rules after-the-fact to evaluate trajectories for compliance with the ROTRs.	An offline-verification tool forms a closed loop with AV development to verify that the AV does not violate a formal rule to maintain sufficient clearance with bicyclists.
Analysis	Analysis of logged system states and identification of formal rule violations can facilitate root cause analysis of problems and help understand the AV's decisions.	The AV performs an uncomfortable stopping maneuver before a jaywalker because it prioritized a formal rule for pedestrian safety over passenger comfort.

rule-guided models of autonomous systems will have to specify what the law requires (Prakken, 2017).

In recent years, several academic works have studied the formalization of ROTRs using different variations of programming and formal logics (Arechiga, 2019; Corso and Kochenderfer, 2020; Esterle et al., 2020). These logical formalisms describe the behavior of the AV in machine-interpretable statements using logical and temporal propositions. Temporal logics (Rescher and Urquhart, 2012) is a class of formal logic methodologies that deals with time-qualified propositions. Temporal logics can formulate natural language specifications (for instance, drive below the posted maximum speed limit at all times and eventually come to a full stop within 1 meter of the stop sign when approaching it) precisely and without any ambiguity for machine interpretability. While a temporal logic formula is agnostic to specific implementations of AV software, different interpretations of an ambiguous ROTR will lead to different temporal logical formulas.

Several studies have made efforts to formalize the *German Road Traffic Regulation* using temporal logics. One study encodes ROTRs for overtaking maneuvers in temporal logic formulas, with the purpose of formally specifying legal accountability for AVs (Rizaldi et al., 2017). The authors argue that it would be desirable to clarify ROTR notions such as a "safe distance" through legal and engineering analysis. A recent study formalizes selected ROTRs for driving on interstate highways using a more complex metric-based temporal logic formalism (Maierhofer et al., 2020). The study argues that legal sources and judicial decisions should supplement and concretize ROTRs to bring consistency between the rules for human drivers and the formalized rules for AVs. In the United States, a study (Hekmatnejad et al., 2019) translates the Responsibility-Sensitive Safety (RSS) model (Shalev-Shwartz et al., 2018) into another variant of temporal logic formulas to formalize the behaviors considered safe under that framework. Another study investigates the formalization of selected ROTRs in the California Department of Motor Vehicle's driver handbook to determine right-of-way in uncontrolled intersections using programming logic (Karimi and Duggirala, 2020). In addition to their application in evaluation of AV behavior with respect to compliance with ROTRs, more recent studies (Cho et al., 2019; Sahin et al., 2020; Xiao et al., 2021) demonstrate the feasibility of using formal rules in AV control and real-time decision-making.

While these efforts proceed, it remains challenging to design an AV that exhaustively and explicitly complies with ROTRs. Both developers and policymakers recognize this and currently address the gap through a variety of mitigating mechanisms. In addition to employing a best effort strategy during system development, developers often work closely with local governments and law enforcement to exchange information, knowledge, and data about AV systems and driving protocols (Goodison et al., 2020). While these developments arguably leave a pathway to deploy AVs with a good faith attempt to comply with ROTRs, the lack of accepted specifications of ROTRs as formal rules creates risk since interpretations may vary widely. For example, the city of San Francisco and one AV developer recently disagreed on the legality of an AV taxi stopping for passenger pickup and drop-off in certain locations (Dave, 2021). Today, ROTRs are interpreted subjectively by both human drivers and AV developers. If policy and engineering efforts can converge on more rigorous and specific interpretations of ROTRs, the resulting better alignment could lead to safer and more efficient road transportation system.

Current industry efforts: Rulebooks, KoPilot, and KoSim. Rulebooks, KoPilot, and KoSim are ongoing industry-based efforts that involve developing products based on formal rules as machine-readable versions of ROTRs.

Rulebooks is an approach created by Motional that develops formal rules specifying good driving behavior from a number of sources (Censi et al., 2019). A Rulebook encodes the formal rules in a priority structure to evaluate preferences among competing trajectories in a given scenario. While maximizing ROTR compliance is a key component of Rulebooks, the framework extends beyond the specification of ROTRs as formal rules in that it aims to formulate a range of behaviors that characterize good driving (Bin-Nun et al., 2020; Collin et al., 2020; Xiao et al., 2021). KoPilot and KoSim are technologies developed by Kontrol for encoding ROTRs into rules and verifying a vehicle's behavior either in simulation (KoSim) or in the real-world (KoPilot) to enable validation of regulatory compliance. The goal of KoPilot in KoSim is to ensure safe and lawful behavior of AVs and enable certification of AVs based on an independent technology (Kontrol, 2018).

Rulebooks and KoPilot are distinct from safety models such as RSS (Shalev-Shwartz et al., 2018), Safety Force Field (SFF) (Nistér

et al., 2019), proposed criticality metrics (Junietz et al., 2018), or the Model Predictive Instantaneous Safety Metric (MPrISM) (Weng et al., 2020), which are methodologies to evaluate the safety of an AV at any particular instant given the state of the world at that moment. One potential use of these safety evaluations is to restrict AVs from entering dangerous states. However, unlike Rulebooks and KoPilot, these efforts do not explicitly seek to achieve compliance with ROTRs.

Impact of formalizing behavior for other road users

Legal requirements, both legislative and regulatory, are one front on the continuous negotiation between multiple road users for priority in traffic (Tennant et al., 2021). Social scientists have long argued that American culture, which includes interpretation, enforcement, and cultural norms surrounding those laws, generally favors motorized road users at the expense of more vulnerable road users (Moeckli et al., 2007). For example, the cultural idea of jaywalkers, created and promoted by automotive lobbies in the 1920s, became enshrined in ROTRs in many states by essentially making the road the domain of motor vehicles (for example, Nevada Revised Statute 484B.297 (Nevada Legislature, 2022b; Norton, 2011). In a similar vein, studies have noted that in some cases, traffic signals favor cars over pedestrians (Levinson, 2018).

As others have argued (Evans et al., 2020; Hulse et al., 2018; Latham and Nattrass, 2019; Pettigrew et al., 2020), the introduction of AVs will likely impact the relative status of and relationship between other types of road users. Below, we discuss some of the ways in which the specifics of AV behavior may affect interactions with other road users.

Controllable behavior. Human driving styles are very heterogeneous (Anesiadou et al., 2021; Makridis et al., 2020). Heterogeneity is closely related to the flexibility and discretion drivers use to respond to uncommon situations and engage in a give and take with other road users. However, the flip side of driving style heterogeneity is that other road users must account for the fact that a given driver's style, and therefore their future actions, is unknown.

Interaction with AVs may be substantially different. AVs are often designed through a scenario-centered approach where behavior is specified in a variety of traffic scenarios (e.g., yielding to pedestrians in a crosswalk; turning right at an unprotected intersection) and the system is developed and tested to execute the desired behavior in those scenarios (IEEE P2846, 2022; Thorn et al., 2018; Winner et al., 2019). An AV designed to exhibit highly specific and defined behaviors may well execute the same strategy each time it encounters a specific scenario. This behavior might be replicated across every vehicle developed by the same company; taken further, industry standardization could lead to similar behaviors across all AV fleets.

The implementation of designed behaviors may increase the predictability of AVs in many scenarios. While the complexity of real-world traffic scenarios and the possibility of perception or other technical failures means that AV behavior is unlikely to be perfectly predictable, it is possible that other road users will be able to better anticipate how AVs will behave in a given situation.

Predictability can have positive impacts. Considerable research has shown throughput, safety, and energy improvements emerging from coordination of vehicle behavior (although coordination is usually envisioned through vehicle-to-vehicle communication rather than through implementing specific, predictable driving styles) (Olia et al., 2016; Taiebat et al., 2018a). Consistent driving can also give other road users confidence to act when they predict the AV will yield precedence (e.g., if pedestrians can be confident that the AV will yield at a crosswalk, then they may be more likely to be assertive). Predictability can, ironically, have unpredictable impacts because it naturally directs other road users to the limits of the AVs permissions (e.g., other road users may learn to increasingly take precedence when negotiating with an AV). Some research has already focused on the possibility that building in hard constraints on AV behavior may lead to unstable outcomes in AV-pedestrian interactions (Fox et al., 2018).

The controllability of AV behavior also implies the possibility of place and culture-specific behavior. For example, AVs could be programmed to be more deferential—or more assertive—in areas with dense pedestrian traffic. AVs could be designed to operate in certain specific environments and optimize their operational characteristics for those environments (Bin-Nun and Binamira, 2020). If AV behavior were made similarly modular, one could imagine behavior that better fits the risk profile and driving characteristics of specific locations (Bin-Nun, 2021). Developers might also choose to tune behavior for any of a wide range of reasons, which might include business-related factors. Therefore, the ability to modulate behavior across time, space, and operating conditions only raises the stakes for the decision and stakeholder input process for designing behavior (Reed et al., 2021).

Harden behaviors. Studies have already pointed out the possibility that requiring AVs to follow behavioral rules, including ROTRs, will "harden" behaviors by aligning AVs with certain desired behaviors (Tennant et al., 2021). This represents a general limitation of most rule-based decision systems; humans naturally have large sets of decision criteria and can consider highly complex interplay of multiple factors in making decisions (Latham and Nattrass, 2019; Suchman and Weber, 2016).

Purely rule-based systems cannot anticipate every potential combination of circumstances. Therefore, behavior specifications imposed as hard rules (e.g., always behave a certain way or maintain a certain distance as a safety margin) have the potential to lead to less nuanced, responsive driving (Xiao et al., 2021). Codifying hard behavioral constraints can create a reality in which the vehicle chooses to behave a certain way to satisfy rules even if there are reasonable considerations for a different course of action. Even if it includes a priority structure with all rules that matter in different contexts, a rule-based system will not have the same degree of leeway as human drivers typically afford themselves. Note that the same will likely be true for machine-learned driving systems, as long as they are held to some set of hard behavioral constraints. Moreover, as with rule-based systems, machine-learned systems will also be limited by the scenarios they have been trained on (Grosan and Abraham, 2011).

The impact of rule-based behavior on other road users will have strong dependence on what constraints are encoded. In many cases, ROTRs are written in a way that is far more deferential to vulnerable road users (VRUs) than actual practice (Schneider and Sanders, 2015). If AVs were to follow a behavior specification that is more deferential than most drivers, it could lead to greater priority for VRUs and shift the hierarchy of users towards non-motorized road users. On the other hand, codification of behavior could easily end up reflecting the current hierarchy and further cementing it. If that were to occur, AV deployment could solidify the current order of priorities on the road and make it even more difficult to change the culture on public roads.

Stakes of formalization. The impacts of AV behavior may go beyond its riders to the rest of the transportation system. If AVs gain market share and represent a significant fraction of traffic in

the area, their behavioral patterns will likely impact mobility for other road users. With a large enough presence in a community, AVs are likely to either alter or amplify the existing culture. Therefore, all road users might be considered stakeholders in how AVs behave and may wish to try to impact expectations for AV behavior.

Even if AVs are not widespread, the very process of formalizing behavior may serve as a forum where stakeholders compete for primacy on the roads. To the extent that public institutions are involved in setting AV behavior, this can be seen as a contest for the cultural definition of proper driving. Much is at stake—some actors may wish to forward a vision for driving behavior that is more centered around non-motorized transportation, while others would like behavior to prioritize the efficiency and throughput of motorized transportation. In many ways, this could be a replay of the contests around defining proper behavior for pedestrians on the roads in the 1920s (Norton, 2011); AVs would be an important vector for defining the local driving culture. Since behavior can be specific to a place, if there were regulatory or other public processes for defining location-specific behavior, this could lead to the emergence of highly differentiated driving cultures in different locales.

This raises the importance of any public processes that could provide input to the definition of AV behavior. As noted currently, regulatory attempts to define AV behavior are nascent and mostly limited to requiring consideration of local ROTRs. However, as some have already called for government involvement in setting digital rules of the road or using a public process to define ethical goal functions (Reed et al., 2021), those processes could end up being perceived as having a significant impact on both AV and human driver operation. They would then be subject to the same competitive forces as current regulatory processes, where private stakeholders frequently invest considerable time and resources to influence (Dal Bó, 2006). Since AV behavior is a complex topic at the cutting edge of technological development, there may be obstacles for non-industry actors to effectively argue for specific behaviors (as they may not be able to convincingly argue for the feasibility or cost of certain behaviors, or understand the broader system implications of requiring certain behaviors).

Note on technological feasibility and development needs. The ability to support a broad stakeholder conversation about the goals and implications of various AV driving styles presupposes a space for having such a conversation. Most of the implications discussed in this section presume that AV behavior can be readily brought in line with external expectations, tuned from location to location, and that the desired behavior is highly modular (e.g., that the prescribed behavior in one scenario is independent from the behavior specified in another).

However, it should be recognized that the creation of a holistic system that would support a consciously directed evolution of driving behavior may require additional effort or development. Our literature review covered a number of commercial and academic endeavors for developing and implementing these capabilities.

A study on the interpretation and formalization of rules of the road

The previous sections discuss policy and engineering considerations for aligning AV behavior with ROTRs and the role of specifying formal rules to achieve such alignment. This section reports insights from a study we conducted to gain insights about possible processes and methods for deriving such formal rules from ROTRs.

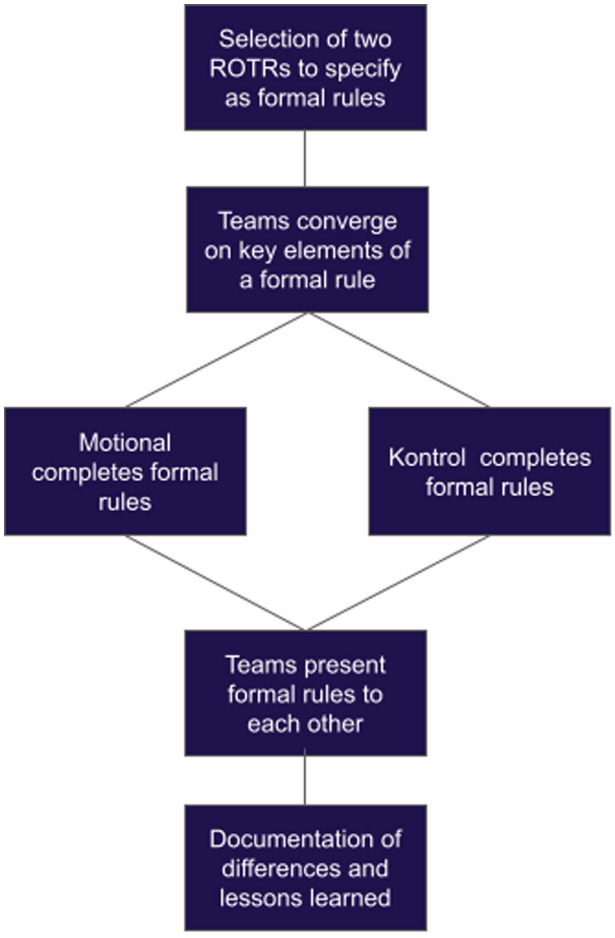

Fig. 1 The study setup. A description of how Motional and Kontrol conducted the traffic law study.

Study setup. The study involved formalizing two ROTRs of the State of Nevada in the United States, where Motional operates an AV service (Motional, 2021). We selected the rules to create a contrast between a rule that involves greater subjective judgment and one that had a clearer numerical specification. Each team worked independently to formalize the two ROTRs (see Fig. 1). To guide the independent work, the teams agreed on a formal rule specification template that includes the following set of elements:

Rule intent and source. A description of the safety, mobility, legal, or other goal the formal rule intends to accomplish. The description includes the basis for the formal rule, which in this study is the corresponding ROTR. For example, the rule intent for a formal rule to stay below the maximum speed limit might be "to comply with the legal maximum speed posted on a road segment." The rule source would be the relevant ROTR.

Rule scope. A set of conditions under which a formal rule applies and rule satisfaction is necessary. For a rule to stay below the maximum speed limit, the rule scope might be any road that has a legal speed limit.

Rule formulation. A logical statement that specifies when a formal rule is violated or satisfied. The rule formulation may include a violation metric that quantifies the degree of violation of the formal rule when the statement is not satisfied, allowing the AV to minimize violation in the event that it cannot fully satisfy a ROTR.

Table 2 Selected ROTRs for the study (verbatim text from Chapter 484B—Rules of the Road (Nevada Legislature, 2022b)).

Short name	ROTR	Key text[a]
Yielding	"484B.250 Vehicle approaching or entering intersection"	"The driver of a vehicle approaching an intersection shall yield the right-of-way to a vehicle, which has entered the intersection from a different highway."
Use of Turn Signals	"484B.413 Requirements for turning on highway; signal for stopping or decreasing speed"	"A signal of intention to turn right or left, or otherwise turn a vehicle from a direct course, shall be given continuously during not less than the last 100 feet traveled in a business or residential district and not less than the last 300 feet traveled in any other area prior to changing the course of a vehicle. This rule shall be observed, regardless of the weather."

[a]The entire text of a ROTR can influence its interpretation. For brevity, we quote the key section of the ROTR that most directly informs the formal rule specification.

For a rule to stay below the maximum speed limit, the rule formulation might be: $v_{ego}(t) \leq v_{max}(t)$ at all times t, where $v_{ego}(t)$ is the speed of the AV at time t, and $v_{max}(t)$ is the posted maximum speed limit on the road segment that the AV travels on at time t. The violation metric might be an increasing function of the excess speed of the AV above the posted maximum speed limit.

Selected rules of the road. Table 2 shows the two State of Nevada ROTRs selected for this study: NRS 484B.250 (Yielding) and NRS 484B.413 (Use of Turn Signals) (Nevada Legislature, 2022b).

The ROTRs present different, but complementary challenges. For Yielding, assessing whether one driver has yielded the right-of-way to another typically involves some judgment. Relevant ROTRs often require drivers to yield the right-of-way in certain situations without specifying how the driver, or law enforcement, would understand whether a given decision is consistent with the obligation to yield. Therefore, a key step in formalizing this ROTR would be to define and formalize a notion of yielding. While mathematical models exist to model when drivers may yield during traffic conflicts, they stop short of presenting a formal definition and specification of what it means to yield (Ma et al., 2017).

The ROTR Use of Turn Signal is *prima facie* more clear-cut in that the ROTR mentions fairly specific parameters and is closely conditioned to physical maneuvers such as turning "from a direct course" (Nevada Legislature, 2022b).

Findings. We found significant overlap in the mathematical formalism the two groups used to express spatial and temporal conditions. However, there were also significant differences in the assumptions, interpretation, and approach used for translating the ROTRs into formal rules.

Motional's approach generally focused on extracting the core intention of the ROTR and crafting a specification that meets both the letter and intention of the legal ROTR. The emphasis on meeting the intention of the written ROTR resulted in broader and more restrictive formal rules than a strict interpretation of the law. This may reflect the Rulebooks approach of combining ROTR compliance with other driving objectives in a general behavior specification.

Kontrol's method adhered as closely to the text of the ROTR as possible to avoid misinterpreting or missing a part of the law. Kontrol translated the text with the understanding that things not explicitly written in the selected sections were covered by other ROTR text. This resulted in very specific rules that narrowly focused on the chosen text only.

Another finding of the study was the inter-dependency between the use case and the rule formulations. Kontrol's main use case for rules is online verification. Therefore, performance considerations influenced the definition of the mathematical framework and, as a result, the translation. Similarly, the translation was influenced by assumptions about the information that is available at run-time (during on-road operation).

While we could go into detail here how the two teams translated the rules and compare the results, we quickly came to realize that there is a lot of room for interpretation in even those two rules. The two interpretations might not be representative of the wide variety of interpretations that might exist in a larger study. We therefore broaden the discussion on findings and instead present, for the two ROTRs, which elements can lead to significant differences in interpretation.

Yielding. The ROTR on yielding refers to an obligation to yield at an intersection. To discuss this ROTR, we first define several concepts. The **yielder** is the vehicle that has to yield the right-of-way. The **yieldee** is the vehicle that has the right-of-way. The origin of the vehicle ("from a different highway"), as well as the location of the intersection and the temporal relationship between the vehicle trajectories ("has entered the intersection") determine which vehicle is the yielder and which is the yieldee. The conflict section is the area that the trajectories of two vehicles share. Figure 2 illustrates the conflict section using an example of two vehicles (A and B) and their trajectories, represented by vehicle outlines at time steps t, with $t_{10} > t_1$, and $t'_{10} > t'_1$.

The challenge here is the determination of yielder and yieldee. What if two vehicles are approaching an intersection from different highways and at very different speeds? What if two vehicles approach the intersection at the same time?

Another source leading to potential differences in rule interpretations is the definition of the conflict section. A strict interpretation could define the entire intersection as the conflict zone, requiring that the yielder not enter the intersection before the yieldee has cleared it. A more lenient interpretation can reduce the size of the conflict zone to a much smaller area.

Studies on traffic conflicts (Hydén, 1987) and post encroachment time (Allen et al., 1978; Archer and Young, 2010) have computed this conflict section using spatial and temporal information. Formalizing a notion of yielding using these concepts may involve prediction algorithms to predict the future path of at least one vehicle and parameters to specify the necessary spatial and temporal distance between yielder and yieldee. Path (or trajectory) prediction can be complex and is, to date, a highly active field of research. There are no standardized methods available, and many companies develop their own, proprietary solutions.

The determination of rule compliance therefore depends on various factors, including the determination of who has the right-of-way in a given scenario, the parameters that define the size of the conflict section, and, in some applications, the prediction mechanism to compute future trajectories for vehicles. Differences in choices for any of these mechanisms or parameters might lead to a different evaluation of rule compliance, where one

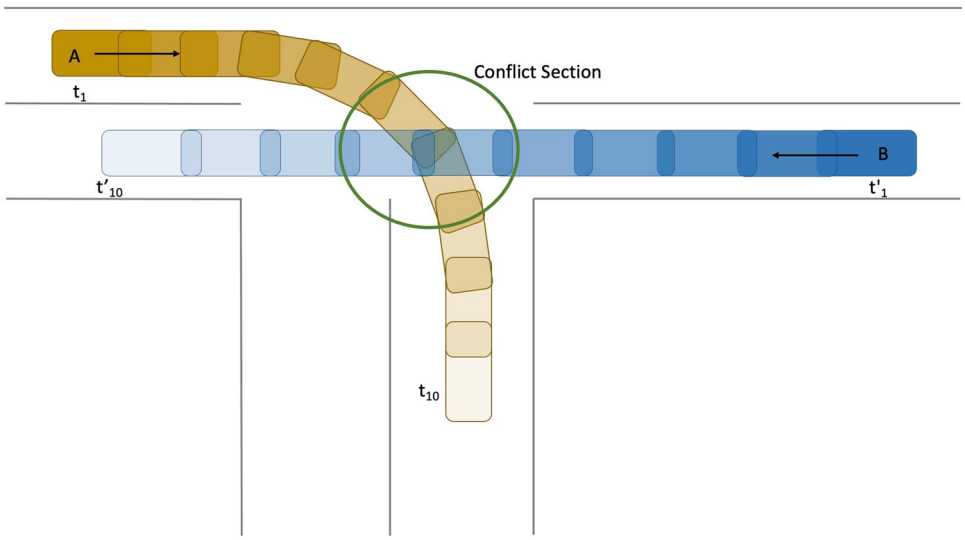

Fig. 2 Illustration of yielding scenario. This figure shows the trajectories of two vehicles approaching an intersection where one vehicle is required to yield the right-of-way to another.

approach might determine a rule violation for a given trajectory in a given scenario while another approach does not.

Given the absence of a clear definition of yielding in the corresponding ROTR text, there was significant interest in exploring other bases for selecting parameters. A promising avenue emerges from the study of the road safety literature, which tries to characterize the risk of situations invoking yielding behavior (e.g., Paul, 2019). Section "Challenges and recommendations" will discuss the potential to integrate external concepts of safety into formal rules.

Use of turn signals. To illustrate the complexity of translating this law, we analyze the various elements in the text and discuss how they can lead to different interpretations.

"*A signal of intention ...*"

We interpreted this as the use of turn signals. Additional ROTRs (such as NRS 484B.420) describe the use of hand signals in case turn signals are not operational. Such laws are relevant for human drivers, but may not be applicable for AVs. Instead, AVs might contain a mechanism to check whether turn signals are operational, which is a precondition for being able to evaluate a formal rule derived from this ROTR. Although not explicitly stated, the ROTR implies that the direction of the turn signal corresponds with the direction of the turn, which the formal rule would need to encode.

"*... to turn right or left, or otherwise turn a vehicle from a direct course, ...* "

This literal description does not preclude swerving or driving on a curved road as turning, although most likely that would be a misinterpretation of the intent of the ROTR. The beginning of a turn needs further definition for identification, for example through a lane marker at an intersection. When using a complete trajectory for rule evaluation, one can compare the direction of the road with the path of the vehicle.

"*... shall be given continuously during not less than the last 100 feet traveled in a business or residential district and not less than the last 300 feet traveled in any other area prior to changing the course of a vehicle....*"

Interpreting this law highlighted how different implementations and use cases can significantly impact the feasibility of complying with the ROTR. For example, one can, with relative ease, verify whether the AV used a turn signal for a sufficient distance when using information from a complete trajectory (by computing the distance between the first time the vehicle starts signaling and the beginning of the turn). However, during online verification (real-time analysis), systems are often designed to only make available a small portion of the trajectory to the verification engine. Therefore, many systems might find it challenging to consider both the beginning and end of a turn signal event where the signal remained on for a significant amount of time. In some situations, an AV system may divert from previously planned trajectories during the course of a maneuver, making it possible to identify a rule violation only in hindsight.

This ROTR also illustrates the importance of providing the system with the correct contextual information (e.g., whether the AV is in a business or residential district). The need for this contextual information may influence system requirements for the map data or the perception system.

The ROTR does not specify the maximum signaling distance, thus Kontrol's literal translation did not capture such a distance, assuming that such a rule is captured in a different ROTR. Motional, however, derived a maximum signaling distance based on other sources and included it into the formal rule for this ROTR.

"This rule shall be observed, regardless of the weather."

While this addition might be of interest to human readers, it does not change the meaning of the previous descriptions and thus does not seem to provide information necessary for the development of a formal rule.

While the ROTR specifies the minimum signaling distance, it does not consider the possibility of a vehicle traveling on a road for less than 100 or 300 feet before making a turn. In such a case, a turn that complies with this ROTR is not possible. One can readily construct cases in which compliance with this ROTR would lead to undesired difficulties in navigating common scenarios (e.g., not being able to make a turn at the end of a short block that a vehicle turned onto; not being able to take an entrance ramp to a highway if one needs to make a turn shortly before getting to the ramp).

The minimum turn signal distance in Nevada ROTR 484B.413 can be interpreted as being in conflict with Nevada ROTR 484B.223. Nevada ROTR 484B.223 says that "a vehicle must not travel more

Table 3 Observations related to the development of formal rules from ROTRs.

Formalization of a ROTR	There is a trade-off between translating the text as stated and capturing the intended driving concept behind the ROTR. While staying close to the text might make it easier to prove that a formal rule covers a ROTR, many details are often not explicitly stated and must therefore be added by interpreting the intention.
Level of abstraction	Formal rules may serve different objectives in different applications. In some online verification applications of systems with small prediction and planning horizons, formal rules may attempt to specify the desired behavior in the context of imperfect information from the system. Other applications may aim to formalize rules to specify the desired driving behavior independent of the capabilities and limitations of the system being evaluated so that they can be applied to any system. The two applications could result in different formal rules.
Interpretation of a violation	A rule violation can indicate • A violation of a ROTRs by the AV. • A scenario in which it is impossible to follow all ROTRs, pointing to a possible inconsistency in the ROTRs. Depending on the application, a rule violation can lead to decisions ranging from simple logging of the event to altering driving behavior in real-time if rule verification is applied online. To aid in the evaluation or decision process, a priority structure among the formal rules may be necessary.
Dealing with ambiguities in the ROTR text	Many different sources can help to resolve ambiguities, ranging from common sense to existing studies or analysis of human driver data. This study suggests that it is unlikely that two independent parties will consistently resolve ambiguities the exact same way.

than 200 feet in a center turn lane before making a left-hand turn from the highway" (Nevada Legislature, 2022b). If the center turn lane (also known as suicide lane) is outside a business or residential district, then the minimum distance for signaling (300 feet) and the maximum distance for turning (200 feet) are in conflict. In this interpretation, entering the center turn lane is considered separate from performing the left turn. While it may be possible to comply with the correct signaling distance before entering the turn lane, the maximum signaling distance for performing the left turn after entering the turn lane is bounded by the maximum distance a vehicle is allowed to travel in the center turn lane.

Study summary. Table 3 summarizes some of the decisions to make when specifying ROTRs as formal rules, and how formal rules may differ.

Looking at two distinct ROTRs highlighted the range of challenges in translating ROTRs into formal rules. In the case of yielding, formalizing the undefined notion of yielding itself was the core challenge. In the case of the apparently more straightforward use of turn signals, challenges emerged from different possible interpretation of the written ROTR.

Challenges and recommendations

Challenges in formalizing rules of the road. Formalizing ROTRs as well-defined, mathematical rules could lead to significant benefits. Formal ROTRs could allow AVs to be designed to follow those rules to the greatest extent possible, which, in turn, has the potential to enable safer and more consistent driving. An AV that follows rules will likely be a more predictable road user for other drivers, especially if those rules are explicitly disclosed. The existence of these rules also might allow for different cultures and localities to specify behavior for AVs, which could promote integration into the local driving culture. Additionally, the creation of a single source of truth for what is considered good driving would allow the synchronization of behavior across AV developers and could potentially contribute towards a safer road transportation system.

However, the literature review, policy analysis, and study highlighted several important obstacles to translating ROTRs into formal behavioral specifications.

First, ROTRs are written by humans, obeyed by humans, enforced and adjudicated by humans, and are embedded in a legal and social context that has interests beyond good driving (Woods, 2021). The ROTRs examined here, like many other ROTRs, are qualitative and make considerable and frequent appeals to judgment. Cultural and regional norms and understandings may influence how rules are interpreted. Therefore, multiple interpretations of the same ROTR are possible, and there currently is no clear process for deciding a priori what behavior is legal.

Secondly, even if each ROTR was written in a fully mathematical form, this would not be sufficient to fully determine behavior. As the Law Commissions of the UK and Scotland and others have noted, ROTRs can conflict or give incompatible guidance for a particular situation (Motional, 2021; Prakken, 2017; Scottish Law Commission, 2018). A driver navigating urban driving may face a choice between complying with some subset of rules and violating another subset—a topic on which the legal frameworks give little guidance. Since ROTRs generally do not include a description of relative priority with other rules, a full behavioral specification is necessary to resolve these conflicts.

Finally, ROTRs themselves benchmark behavior against external notions of safety. For example, the Nevada ROTRs express that the duty of a driver to yield the right-of-way when entering a highway extends "until the driver may proceed with reasonable safety" (Nevada Legislature, 2022b). The fact that a ROTR references safety as a determinant of legal behavior suggests that there is a notion of safety that is external to the behavior specified in the ROTR. To fully codify the behavior in this rule, a developer would need to separately create a conception of safety to specify when proceeding onto the highway is allowed under the rule. Given that not all AV systems have the same capabilities, what is safe for a more capable system is not necessarily safe for a more limited system. Yet, many would argue that all road users should follow the same rules when interacting with other road users. This tension adds another layer of complexity to creating consensus on interpreting ROTRs.

Today, these questions are largely left for AV developers to answer individually, with some incremental aspects of these broad questions addressed collaboratively through activity in standards and regulatory forums. However, while the technical work to implement behavior might well be considered an appropriate arena of competition for the AV industry, the definition of what represents acceptable driving on public roads is inherently a matter of broader societal interest. The stakeholders include other road users, law enforcement, and the public at large. Therefore, difficulties in extracting a definition of driving behavior from legal documents might be seen more as a gap in public policy than as a challenge for developers. We suggest mechanism for addressing this gap in the remainder of this section.

Research recommendations. The previous sections raised several obstacles to extracting behavioral specifications from ROTRs. The reality that ROTRs contain significant ambiguity has long been recognized, including outside the context of the AV industry (Rothengatter, 1997; Woods, 2021). There are already numerous rationales for better drafting of ROTRs to remove elements of subjectivity; the public interest in predictable and synchronized AV behavior adds to this list. We anticipate political challenges as the distinct policy-making centers that regulate on-road behavior and vehicle design come into greater contact.

Dealing with apparent conflicts between ROTRs is an emerging focus of research (Censi et al., 2019). This article earlier referenced the concept of a rule violation metric; the need for such a metric emerges from an interest in describing violations of different ROTRs using a common violation metric. Violation metrics can be leveraged to trade-off violations of one rule for another when necessary.

We have identified the necessity of a concept of safety external to ROTRs to determine rule compliance. There is considerable ongoing work in government, industry, and academia to assess the safety of a given driving situation. These are ripe candidates for further development into a safety concept. Within the context of the ROTR framework, it may be possible to delve deeper into the case law and precedents involving ROTRs, which can shed light on how driving rules are interpreted. While it seems unlikely that examining judicial records will allow for convergence on a single interpretation of ROTRs, it should be seen as one strategy among many to better derive behavioral specifications.

Policy recommendations and conclusions

This article has addressed a broad range of questions at the intersection of engineering, policy, and safety for AVs. Unlike human drivers, AVs hold the prospect of implementing carefully designed behavior, which represents an opportunity for greater societal input into their driving decisions. We have explored the potential implications of AV driving behaviors for other road users and how the deployment of AVs presents an opportunity to either modify or harden existing relationships between different road users. Finally, we have shown that ROTRs offer, at best, only a partial answer to this question, and may not be adequate as an answer to the question of "how does society believe AVs should drive?"

These findings speak to a need for both political and technical advances in specifying driving behavior for AVs. The political process by which ROTRs are generated and enforced do not currently integrate well either with the development of AVs or the standards and regulations that govern AV development. As vehicle automation becomes responsible for more driving, the current legal framework for governing driving behavior (i.e., ROTRs) will likely become less important as a tool for ensuring safety. Policymakers at all levels should actively consider which institutions, whether at the local, national, or international levels, should govern driving behavior on the road, and what processes will create the detailed and specific guidance that can align behavior across disparate AV developers and road actors. Regulators might also consider, given the issues identified in requiring AVs to comply with ROTRs, to adopt a phased approach where responsibility to comply with ROTRs grows over time or as an AV fleet scales from testing to broader deployment.

The choices civil society and regulators make to govern AV behavior will likely reverberate well beyond AVs and may impact how road users consider the space of public roads. The very process of designing AV behavior may force society to once again grapple with the broader question of who our roads are for and what values should govern behavior on those roads.

This paper detailed current challenges in creating a comprehensive behavioral specification as well as ongoing approaches to address the identified gaps. The previous subsection outlines a research roadmap towards more comprehensive behavioral specifications, including the integration of ROTRs into behavioral specifications. There is a strong case that the technical research agenda cannot be separated from the political interests in this research. Several factors suggest that research on this topic should be performed collaboratively across industry and other sectors: Behavioral specifications are a matter of significant public interest and can be technology agnostic (the right driving behavior is independent of whether the driver is a human or an AV or how an AV is built). Along these lines, the Law Commissions of the UK and Scotland recommended establishing a forum to better align industry interpretations of ROTRs. Ideally, this would be not just a technical forum to mathematically capture ROTRs, but a forum to capture stakeholder input as to what values should be reflected in driving. Engaging industry stakeholders in this forum and similar ones will likely require political effort and prioritization to succeed.

Progress on both technical frameworks and political governance of driving behavior would result in better, more comprehensive behavioral specifications for AVs. More research into driving behavior could also democratize input to the conversation on driving behavior by making this technical topic more accessible to a broader range of stakeholders. Improving driving behavior is one of the most important pathways towards improving the safety of our roadways. Aligning the political process for defining good driving behavior with the technical progress necessary to implement that behavior on an AV would likely serve as an important tool for progress on roadway safety.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Received: 29 October 2021; Accepted: 2 August 2022;
Published online: 30 August 2022

References

Allen BL, Shin BT, Cooper PJ (1978) Analysis of traffic conflicts and collisions. Transp Res Board (667):67–74

Alliance for Automotive Innovation (2022) Alliance for automotive innovation, policy roadmap to advance automated vehicle innovation. https://www.autosinnovate.org/innovation/AVRoadmap.pdf. Accessed Jun 2022

Anesiadou A, Makridis M, Mattas K, Fontaras G, Ciuffo B (2021) Characterization of drivers heterogeneity and its integration within traffic simulation. https://arxiv.org/abs/2107.02618

Archer J, Young W (2010) The measurement and modelling of proximal safety measures. In: Proceedings of the Institution of Civil Engineers-Transport vol 163, pp. 191–201. Thomas Telford Ltd

Arechiga N (2019) Specifying safety of autonomous vehicles in signal temporal logic. In: 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019. pp. 58–63

Automated Vehicle Safety Consortium (2022). Automated vehicle safety consortium. https://avsc.sae-itc.org/. Accessed Jun 2022

Bin-Nun AY, Binamira I (2020) A framework for the impact of highly automated vehicles with limited operational design domains. Transp Res Part A Policy Pract 139:174–188

Bin-Nun AY, Panasci A, Tebbens RJD (2020). Heinrich's triangle heavy-tailed distributions and autonomous vehicle safety. In: Transportation Research Board Annual Meeting, Washington

Bin-Nun AY (2021) System and method for improving autonomous vehicle safety performance, Sept 7 2021. US Patent 11,112,797

Blais E, Dupont B (2005) Assessing the capability of intensive police programmes to prevent severe road accidents: A systematic review. British Journal of Criminology 45(6):914–937

Brown A, Rodriguez G, Hoang T, Safford H, Anderson G, Cohen D'Agostino M (2018) Federal, state, and local governance of automated vehicles. Institute of Transportation Studies & Policy Institute for Energy, Environment and the Economy, University of California Davis, 2018

Brunello A, Montanari A, Reynolds M (2019) Synthesis of LTL formulas from natural language texts: state of the art and research directions. In : 26th International Symposium on Temporal Representation and Reasoning (TIME 2019). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik

Censi A, Slutsky K, Wongpiromsarn T, Yershov D, Pendleton S, Fu J, Frazzoli E (2019) Liability, ethics, and culture-aware behavior specification using rulebooks. In: 2019 International Conference on Robotics and Automation (ICRA). IEEE, pp. 8536–8542

Chakraborty D, Chaisse J, Pahari S (2020) Global auto industry and product standards: A critical review of India's economic and regulatory experience. J. Int Trade Law Policy 19(1):8–35

Channon M, McCormick L, Noussia K (2019) The law and autonomous vehicles. Taylor & Francis

Cho K, Ha T, Lee G, Oh S (2019) Deep predictive autonomous driving using multi-agent joint trajectory prediction and traffic rules. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2076–2081. https://doi.org/10.1109/IROS40897.2019.8967708

Claypool H, Bin-Nun AY, Gerlach J (2017) Self-driving cars: the impact on people with disabilities. Ruderman Family Foundation, Newton, MA, USA

Collin A, Bilka A, Pendleton S, Radboud Duintjer, Tebbens R (2020) Safety of the intended driving behavior using rulebooks. In: 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE. pp. 136–143

Corso A, Kochenderfer MJ (2020) Interpretable safety validation for autonomous vehicles. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1–6

Dal Bó E (2006) Regulatory capture: a review. Oxford review of economic policy 22(2):203–225

Dave P (2021) GM's cruise disputes San Francisco concerns on stops, says 'double parking' legal, Dec 2021. URL https://www.reuters.com/business/autos-transportation/gms-cruise-disputes-san-francisco-concerns-stops-says-double-parking-legal-2021-12-07/

De Freitas J, Censi A, Smith WB, Lillo LD, Anthony S. E, Frazzoli S (2021) From driverless dilemmas to more practical commonsense tests for automated vehicles. Proc Natl Acad Sci USA 118(11):e2010202118

Esterle K, Gressenbuch L, Knoll A (2020) Formalizing traffic rules for machine interpretability. https://arxiv.org/abs/2007.00330

Evans K, de Moura N, Chauvier Stéphane, Chatila R, Dogan E (2020) Ethical decision making in autonomous vehicles: the AV ethics project. Sci Eng Ethics 26(6):3285–3312

Federal Minister for Transport, Innovation and Technology (2019) Verordnung des Bundesministers für Verkehr, Innovation und Technologie, mit der die Automatisiertes Fahren Verordnung geändert wird (1. Novelle zur AutomatFahrV). https://www.ris.bka.gv.at/Dokumente/BgblAuth/BGBLA_2019_II_66/BGBLA_2019_II_66.pdf. Accessed: Jun 2022

Federal Ministry of Transport and Digital Infrastructure (2021) Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes—Gesetz zum autonomen Fahren. http://www.bmvi.de/SharedDocs/DE/Anlage/Gesetze/Gesetze-19/gesetz-aenderung-strassenverkehrsgesetz-pflichtversicherungsgesetz-autonomes-fahren.pdf. Accessed: Jun 2022

Feen G, Bin-Nun AY, Panasci A (2020) Fostering economic opportunities through autonomous vehicle technology. Securing America's Future Energy

Five AI (2019) Towards a digital highway code. https://eprints.whiterose.ac.uk/160312/1/fiveai_cert_paper_v1.pdf. Accessed: Jun 2022

Fox C, Camara F, Markkula G, Romano R, Madigan R, Merat N et al. (2018) When should the chicken cross the road?: Game theory for autonomous vehicle-human interactions. Proceedings of the 4th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2018), pp. 431–439

Goodison SE, Barnum JD, Vermeer MJD, Woods D, Lloyd-Dotta T, Jackson BA (2020) Autonomous Road Vehicles and Law Enforcement: Identifying High-Priority Needs for Law Enforcement Interactions With Autonomous Vehicles Within the Next Five Years. RAND Corporation, Santa Monica, CA

Grosan C, and Abraham A (2011) Rule-based expert systems. In: Intelligent systems. Springer, pp. 149–185

Haas AK (2001) Chipping away at state tort remedies through pre-emption jurisprudence: Geier v. American Honda Motor Co. Calif Law Rev 89:1927–1950

Hekmatnejad M, Yaghoubi S, Dokhanchi A, Amor HB, Shrivastava A, Karam L, Fainekos G (2019) Encoding and monitoring responsibility sensitive safety rules for automated vehicles in signal temporal logic. In: Proceedings of the 17th ACM-IEEE International Conference on Formal Methods and Models for System Design, pp. 1–11

Hulse LM, Xie H, Galea ER (2018) Perceptions of autonomous vehicles: relationships with road users, risk, gender and age. Safety Sci 102:1–13

Hydén C (1987) The development of a method for traffic safety evaluation: The swedish traffic conflicts technique. Bulletin. University of Lund, Lund institute of technology, Department of traffic planning and engeneering

IEEE P2846 (2022) IEEE standard for assumptions in safety-related models for automated driving systems. IEEE Vehicular Technology Society, pp. 1–59

ISO (2018). ISO 26262:2018 Road vehicles—Functional safety. URL https://www.iso.org/standard/68383.html

ISO (2019). ISO/PAS 21448:2019 Road vehicles—Safety of the intended functionality

ISO (2020) ISO/TR 4804:2020, Road vehicles—safety and cybersecurity for automated driving systems—design, verification and validation

Junietz P, Bonakdar F, Klamann B, Winner H (2018) Criticality metric for the safety validation of automated driving using model predictive trajectory optimization. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 60–65

Kalra N, Groves DG (2017) The enemy of good: Estimating the cost of waiting for nearly perfect automated vehicles. Rand Corporation

Karimi A, Duggirala SP (2020) Formalizing traffic rules for uncontrolled intersections. In: 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS). IEEE, pp. 41–50

Kate RJ, Wong YW, Mooney RJ (2005) Learning to transform natural to formal languages. In: AAAI, vol. 5, pp. 1062–1068

Kontrol (2018) Kontrol. http://www.kontrol.tech. Accessed: Jun 2022

Koopman P, Ferrell U, Fratrik F, Wagner M (2019) A safety standard approach for fully autonomous vehicles. In: International Conference on Computer Safety, Reliability, and Security. Springer, pp. 326–332

Latham A, Nattrass M (2019) Autonomous vehicles, car-dominated environments, and cycling: Using an ethnography of infrastructure to reflect on the prospects of a new transportation technology. J Transp Geogr 81:102539

Levinson D (2018). How traffic signals favour cars and discourage walking. URL https://theconversation.com/how-traffic-signals-favour-cars-and-discourage-walking-92675. Accessed: Jun 2022.

Ma Z, Sun J, Wang Y (2017) A two-dimensional simulation model for modelling turning vehicles at mixed-flow intersections. Transp Res Part C Emerg Technol 75:103–119

Maierhofer S, Rettinger A-K, Eva Charlotte, Mayer EC, Althoff M (2020) Formalization of interstate traffic rules in temporal logic. In: 2020 IEEE Intelligent Vehicles Symposium (IV). IEEE, pp. 752–759

Makridis M, Leclercq L, Ciuffo B, Fontaras G, Mattas K (2020) Formalizing the heterogeneity of the vehicle-driver system to reproduce traffic oscillations. Transp Res Part C: Emerg Technol 120:102803

Ministry of Transportation (2022) Ontario's automated vehicle pilot program. http://www.mto.gov.on.ca/english/vehicles/automated-vehicles.shtml. Accessed: Jun 2022

Moeckli J, Lee JD (2007) The making of driving cultures. Imp Traffic Saf Cult US 38(2):185–192

Motional (2021) Motional response: Law commission of england and scottish law commission consultation paper 3. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2021/06/AVRF073-Motional.pdf. Accessed: Jun 2022

Motional (2021) Voluntary safety self-assessment (VSSA), autonomous vehicle safety ecosystem. https://motional.com/news/safety-the-driver-at-motional

National Conference of Commissioners on Uniform State Laws (2019) Uniform automated operation of vehicles act, national conference of commissioners on uniform state laws, annual conference meeting in its one-hundred-and-twenty-eighth year Anchorage, Alaska. https://www.uniformlaws.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=2dd86096-2546-dfe8-eeb6-91c11e0e1b2. Accessed: Jun 2022

N., Legislature (2022a) Chapter 484b—autonomous vehicles. https://www.leg.state.nv.us/nrs/nrs-482a.html, 2022a. Accessed: Jun 2022

N., Legislature (2022b) Chapter 484b—rules of the road. http://www.leg.state.nv.us/NRS/NRS-484B.html. Accessed: Jun 2022

NHTSA (2017) Automated driving systems 2.0: A vision for safety. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf. Accessed: Jun 2022

Nistér D, Lee H-L, Ng J, Wang Y (2019) The safety force field. NVIDIA White Paper

Norton PD (2011) Fighting traffic: the dawn of the motor age in the American city. Mit Press

Olia A, Abdelgawad H, Abdulhai B, Razavi SN (2016) Assessing the potential impacts of connected vehicles: mobility, environmental, and safety perspectives. J Intell Transp Syst 20(3):229–243

Parker D, Reason JT, Manstead AntonySR, Stradling SG (1995) Driving errors, driving violations and accident involvement. Ergonomics 38(5):1036–1048

Paul M (2019) Safety assessment at unsignalized intersections using post-encroachment time's threshold-a sustainable solution for developing countries. In: Advances in Transportation Engineering. Springer, pp. 117–131

Pettigrew S, Nelson JD, Norman R (2020) Autonomous vehicles and cycling: policy implications and management issues. Transp Res Interdiscip Perspect 7:100188
Prakken H (2017) On the problem of making autonomous vehicles conform to traffic law. Artif Intell Law 25(3):341–363
Reed N, Leiman T, Palade P et al. (2021) Ethics of automated vehicles: breaking traffic rules for road safety. Ethics Inf Technol 23:777–789. https://doi.org/10.1007/s10676-021-09614-x
Rescher N, Urquhart A (2012) Temporal Logic. LEP Library of Exact Philosophy. Springer Vienna. ISBN 9783709176641
Rizaldi A, Keinholz J, Huber M, Feldle J, Immler F, Althoff M, Hilgendorf E, T, Nipkow T (2017) Formalising and monitoring traffic rules for autonomous vehicles in isabelle/hol. In: International conference on integrated formal methods. Springer.pp. 50–66
Rothengatter T (1997) Psychological aspects of road user behaviour. Appl Psychol 46(3):223–234
SAE International (2018) Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles j3016_201806. https://www.sae.org/standards/content/j3016_201806/. Accessed: Jun 2022
Sahin YE R, Quirynen R, Di Cairano S (2020) Autonomous vehicle decision-making and monitoring based on signal temporal logic and mixed-integer programming. In: 2020 American Control Conference (ACC). IEEE, pp. 454–459
Schmitt S (2020) Right of way: Race, class, and the silent epidemic of pedestrian deaths in America. Island Press
Schneider RJ, Sanders RL (2015) Pedestrian safety practitioners' perspectives of driver yielding behavior across north america. Transp Res Rec 2519(1):39–50
Scottish Law Commission (2018) Scottish law commission, automated vehicles, a joint preliminary consultation paper. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2018/11/6.5066_LC_AV-Consultation-Paper-5-November_061118_WEB-1.pdf. Accessed Jun 2022
Scottish Law Commission (2020) Scottish law commission, automated vehicles: consultation paper 3—a regulatory framework for automated vehicles, a joint consultation paper. https://s3-eu-west-2.amazonaws.com/lawcom-prod-storage-11jsxou24uy7q/uploads/2021/01/AV-CP3.pdf. Accessed: Jun 2022
Shalev-Shwartz S, Shammah S, Shashua A (2018) On a formal model of safe and scalable self-driving cars. https://arxiv.org/abs/1708.06374
Singapore Standards Council (2019) Tr 68, technical reference autonomous vehicles. http://www.singaporestandardseshop.sg/. Accessed: Jun 2022
Singh S (2015) Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Technical report, NHTSA
Smith BW (2013) Automated vehicles are probably legal in the United States.Texas A&M Law Rev. 411
Smith BW (2017) How governments can promote automated driving. New Mexico Law Review. vol. 47, N.M. L. Rev. 99
Suchman L, Weber J (2016) Human-machine autonomies. Autonomous weapons systems: law, ethics, policy. Cambridge University Press. pp. 75–102
Taiebat M, Brown AL, Safford HR, Qu S, Xu M (2018a) A review on energy, environmental, and sustainability implications of connected and automated vehicles. Environ Sci Technol 52(20):11449–11465
Taiebat M, Brown AL, Safford HR, Qu S, Xu M (2018b) A review on energy, environmental, and sustainability implications of connected and automated vehicles. Environ Sci Techno 52(20):11449–11465
Tennant C, Neels C, Parkhurst G, Jones P, Mirza S, Stilgoe J (2021) Code, culture and concrete: Self-driving vehicles and the rules of the road. Front Sustain Cities. https://doi.org/10.3389/frsc.2021.710478
Thorn E, Kimmel SC, Chaka M (2018) A framework for automated driving system testable cases and scenarios (Report No. DOT HS 812 623). Technical report, United States. Department of Transportation. National Highway Traffic Safety Administration
UNECE (2018) Regulation No 79 of the Economic Commission for Europe of the United Nations (UN/ECE)—Uniform provisions concerning the approval of vehicles with regard to steering equipment, TRANS/WP.29/343. http://www.unece.org/trans/main/wp29/wp29wgs/wp29gen/wp29fdocstts.html. Accessed: Jun 2022
United Nations Economic and Social Council (2020) Proposal for a new UN Regulation on uniform provisions concerning the approval of vehicles with regards to Automated Lane Keeping System, ECE/TRANS/WP.29/2020/81. https://undocs.org/ECE/TRANS/WP.29/2020/81. Accessed: Jun 2022
Vellinga NE (Ed.) (2021) Automated driving safety data protocol—Ethical and legal considerations of continual monitoring. (Technical Report ITU-T; No. FGAI4AD-02). International Telecommunication Union. http://handle.itu.int/11.1002/pub/81b9a99a-en
Weng B, Rao SJ, Deosthale E, Schnelle S, Barickman F (2020) Model predictive instantaneous safety metric for evaluation of automated driving systems. https://arxiv.org/abs/2005.09999
Winner H, Lemmer K, Form T, Mazzega J (2019) Pegasus-first steps for the safe introduction of automated driving. In: Road vehicle automation 5, Springer, pp. 185–195
Wood M et al. Safety first for automated driving (2019) Aptiv, Audi, BMW, Baidu, Continental Teves, Daimler, FCA, HERE, Infineon Technologies, Intel, Volkswagen
Woods JB (2021) Traffic without the police. Stanf Law Rev 73(6):1471
Xiao W, Mehdipour N, Collin A, Bin-Nun AY, E., Frazzoli E, Tebbens RD, Belta C (2021) Rule-based optimal control for autonomous driving. In: Proceedings of the ACM/IEEE 12th International Conference on Cyber-Physical Systems (ICCPS '21). Association for Computing Machinery, New York, NY, USA, pp. 143–154. https://doi.org/10.1145/3450267.3450542
Yagil D (2005) Drivers and traffic laws: a review of psychological theories and empirical research. In: Traffic and Transport Psychology. Elsevier, Oxford pp. 487–503

Competing interests
The authors declare no competing interests.

Ethical approval
This article does not contain any studies with human participants performed by any of the authors.

Informed consent
This article does not contain any studies with human participants performed by any of the authors.

Additional information
Correspondence and requests for materials should be addressed to Amitai Y. Bin-Nun.