Negotiating the traffic: Can cognitive science help make autonomous vehicles a reality?

Nick Chater,[1] Jennifer Misyak,[1] Derrick Watson,[2] Nathan Griffiths[3] & Alex Mouzakitis[4]

[1]Behavioural Science Group, Warwick Business School, University of Warwick
[2]Department of Psychology, University of Warwick
[3]Department of Computer Science, University of Warwick
[4]Jaguar Land Rover

Abstract

To drive safely among human drivers, cyclists and pedestrians, autonomous vehicles will need to mimic, or ideally improve upon, human-like driving. Yet driving faces us with difficult problems of joint action: "negotiating" with other users over shared road-space. We argue that autonomous driving provides a test case for computational theories of social interaction, with fundamental implications for the development of autonomous vehicles.

Key words. Negotiation, virtual bargaining, autonomous driving,

Acknowledgments. This work was supported by RCUK/Jaguar Land Rover Grant EPSRC EP/N012380/1. The views expressed are solely those of the authors, not the sponsoring bodies. We thank two anonymous reviewers for their valuable input.

Corresponding author. Nick Chater, Behavioural Science Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK, nick.chater@wbs.ac.uk.

Alan Turing famously challenged future generations to create a machine that would be indistinguishable from a person through the medium of typewritten language. The future of fully autonomous vehicles, into which tens of billions of dollars are being invested globally, appears to depend on the cognitive and computational sciences being able to meet a related challenge: the creation of computer systems that can *drive* in a way that blends seamlessly, and safely, into roads populated with human drivers. Yet solving this problem, so that autonomous vehicles, human motorists, cyclists and pedestrians can negotiate our roads safely, involves addressing fundamental questions at the frontiers of cognitive science. These challenges involve familiar issues in perception and control, but also less obvious, and arguably far more difficult, questions concerning the cognitive foundations of social interaction. Thus "negotiating the traffic," is, we suggest, not merely a figure of speech: it involves a tacit process of negotiation with other road users in a safety critical environment, in real-time, and with low-bandwidth communication.

Human interactions are often so effortless that we are unaware of the complexity of the reasoning that our brains performing—computations that autonomous vehicles will need to emulate. The driving situations in Figure 1 illustrate some of the complexities, even in a manoeuver as simple as moving briefly into the on-coming lane to avoid an obstacle. Drivers are playing a game of "chicken"—one, but not both, should give way, to avoid either collision or deadlock.

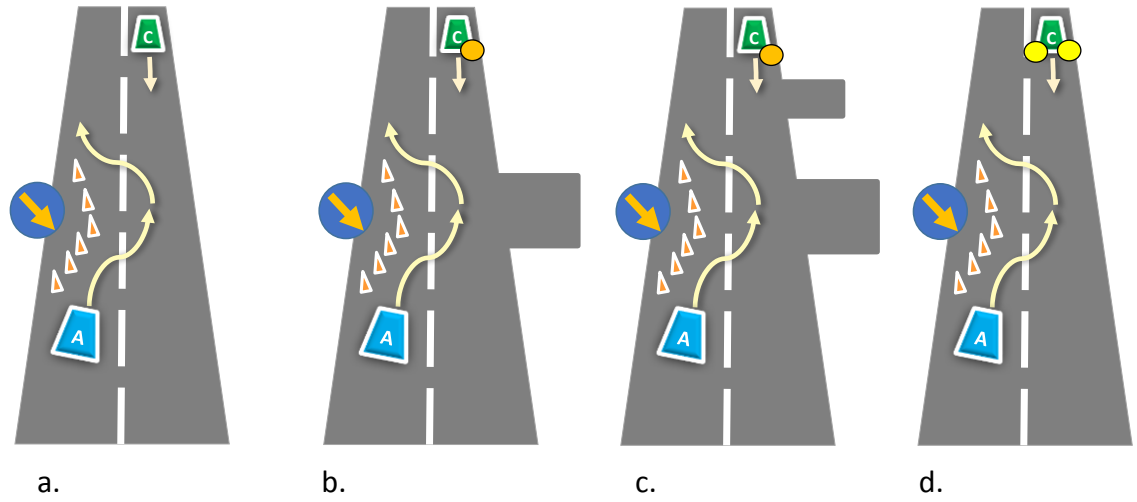

Figure 1. *Who goes first?* In 1a. drivers face a problem of joint action. Car A needs to encroach on the opposite lane to pass an obstruction, but could collide with Car C. The cars face a game of 'chicken'—one must give way; but if both give way, they will fall into deadlock. One natural rule is that priority goes to the driver who stays in lane. But in these scenarios, C is moving slowly, so that, unless C speeds up, A can pass the obstruction successfully. But perhaps C is accelerating? In 1b. C starts indicating left. This would seem to imply an immediate intention to turn left at the point where A would encroach into C's lane; and be tantamount to a "claim" on the "contested" region of road—which implies that A should give way. In 1c., the same indicating signal may now be interpreted as an intention to turn down the small driveway, and hence communicating the opposite message to A, "ceding" the contested area. In 1d., C flashes its headlights, either to yield, or perhaps to signal "I'm coming"—the interpretation may depend on changes in vehicle velocity as well as local informal norms. Many slight variations on this, and similar, scenarios can change the "natural" interpretations of signals and actions of the drivers. An automated vehicle "impersonating" a human driver will be hazardous unless it follows natural human driving behaviour accurately.

Drivers have to engage in a "joint action" (1-4), each aligning their behaviour to that of the other, but without central coordination. One natural, but limited, approach is to assume that each driver treats other drivers as mere "moving objects" whose trajectory is to be predicted, and that each driver optimises their own actions against these predictions. But this type of mutual prediction can lead to vicious circularity: A's prediction about C depends on A's beliefs about C's predictions about A, and so on, indefinitely. Another limited approach is based on systems of rules: e.g., only move into the other lane when no collision will occur if other vehicles maintain their current velocity and path. But as Figure 1 illustrates, the variety of configurations makes creating a closed set of rules which mimics the flexibility of human driving behaviour difficult and perhaps impossible.

How do drivers reach a common implicit agreement? We suggest that agents, whether human drivers or autonomous vehicles, must ask: *if we could negotiate, what agreement would we reach about who does what and when?* It is, of course, crucial that all the agents reach the same agreement, or at least compatible agreements; otherwise deadlock or accident may result.

The problem of such implicit negotiation is at the frontiers of cognitive science for at least three reasons.

(i) Even the explicit negotiation is generally extremely difficult to model, despite decades of intensive analysis in game theory, management and political science, and psychology [5].
(ii) Negotiation can be especially challenging when communication is limited. For example, in Figure 1c, C signals to turn into a driveway, so that A can proceed. But can A be sure this is C's intention? Might C instead be intending to proceed left at the road opposite the obstacle, implying that C is intending to proceed rather than cede the road to A?
(iii) In driving, communication is highly restricted (to signalling, flashing headlights, honking, waving, as well as the 'manner' of vehicle movement, see [6])—and the meaning of such signals is itself likely to require reaching agreement. For example, in Figure 1d, both agents need to know whether C flashing its headlights is an invitation for A to proceed, or a warning that A should stay out of the way (see [7]).

On the other hand, the highly restricted domain of actions and signals, makes the problem of interaction in driving an approachable, if challenging, special case. Our approach to the problem is based on the theory of "virtual bargaining" [8]; we propose that each agent simulates the outcome of a hypothetical bargaining process, based on the common knowledge among agents of their beliefs and goals. Virtual bargaining can be converted into a mathematically precise form using an extension of game theory. But other approaches, e.g., based on team reasoning [9], recursive Bayesian models [10, 11] or lower-level "sensorimotor" communication [12] should also be explored. We suggest that the challenge of understanding and building agents that can genuinely "negotiate" the traffic should be a major focus of cognitive science research, of comparable scale to the major research efforts in computer vision and machine learning that have been focussed on autonomous driving. This will require both experimental work on human driving interactions (whether explored

in abstract lab experiments, driving simulators and real road conditions) and new theoretical developments. It will also requiring creating and testing of agents which, we may hope, can pass the automotive "Turing test" by driving safely and acceptably among other human drivers, initially in software simulations but ultimately, of course, in real driving conditions.

We suggest, moreover, that while the problem of sensing the surroundings and other road users appears to be yielding impressive progress, the challenge of traffic "negotiation" has scarcely been addressed either by the specialist literature on transportation research or by the cognitive sciences [13]. Indeed, the rate of progress on this challenge within the cognitive sciences may prove a decisive limiting factor in the development of autonomous vehicles.

Moreover, the safety of more limited steps to autonomy, where control is handed back and forth to a human driver, may depend on progress on understanding and modelling 'negotiation.' The situations illustrated in Figure 1 arise routinely and unpredictably in urban driving, so such situations ought not to be classified as "too difficult" and handed back to human users. In part, this is because a human will not be able to attend to, and resolve how to act in, such an interaction when previously engaged in some other task; but also because identifying the "difficult" cases which require human-level negotiating skills may not be accurate without the deployment of such skills (just as it is difficult to accurately identify "difficult" chess positions without actually attempting, and struggling, to decide what to do in such positions). Table 1 outlines some of the cognitive science challenges and possible pathways for the development of autonomous vehicles (leaving aside important ethical issues, questions of acceptability of even a small number of accidents, and problems of the opacity of computer algorithms) which have been discussed elsewhere [14, 15]); it includes one scenario in which the challenge of negotiation is addressed and three ways in which it might be skirted.

We believe that the challenge of autonomous vehicles, which promises great gains in human welfare through improved mobility, safety, and environmental impacts, brings to light fundamental challenges for cognitive science and artificial intelligence, not just in sensing and control (where machines may potentially exceed human performance—e.g., in response times), but in mimicking or seamlessly meshing with human behaviour in driving interactions. The problem of understanding how we "negotiate" the traffic also provides a microcosm of deep questions concerning human social interaction and communication more generally.

Driving ecology	**Automated negotiating skills required**	**Open Questions**	**Scale of cognitive science challenges**	**Non-motoring analog**
Unmarked, partially or fully autonomous vehicles mixed with human drivers, cyclists and pedestrians.	Algorithms with human-level negotiating skills.	Can autonomous vehicles safely and acceptably interact with human drivers and other road users?	*Hard*: Create and test a theory of negotiation	The motoring equivalent of passing the Turing Test (presumably much simpler than the original)
Vehicles with publically signalled autonomous and manual modes	Below human-level. Difficult interactions are solved in manual mode by humans, not algorithms.	How easily can other human road-users adjust to interacting with, and predicting, autonomous-model behaviour?	*Medium*: How easily can drivers "pick-up the baton" when shifting out of autonomous mode. How to enhance human-drivers' discrimination of autonomous vs non-autonomous modes?	Autopilot in aircraft—but potentially with much faster and less predictable switching between auto and manual modes.
Highly distinctive autonomous "pods" with highly predictable behaviour and no manual over-ride	Simple, predictable, re-active behaviour	How far do people "anthropomorphize" the artificial drivers as human (e.g., attempting to communicate and negotiate with them).	*Routine*: How can anthropomorphism of the vehicles be minimized? (e.g., by facing passengers away from the direction of travel)	Normal human-machine interaction: the human adjusts to the machine.
Dedicated spaces: separate autonomous vehicles from other road users	None	Is it really feasible to create a fully parallel transport infrastructure?	*Minimal.* Human control is cut out. Human acceptability is the main concern.	Automated subway and rail travel.

Table 1. *Types of autonomous interaction and cognitive science challenges*.

References

1. Sebanz, N., Bekkering, H. & Knoblich, G. (2006). Joint Action: Bodies and Mind Moving Together. *Trends in Cognitive Sciences*, *10*(*2*), 70– 76.
2. Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common ground and coordination in joint activity. In: W. B. Rouse & K. R. Boff (Eds.), *Organizational Simulation* (pp. 139-184). Hoboken, NJ: John Wiley & Sons.
3. Portouli, E., Nathanael, D., & Marmaras, N. (2014). Drivers' communicative interactions: On-road observations and modelling for integration in future automation systems. *Ergonomics, 57,* 1795-1805.
4. Swan, L. A. & Owens, M. B. (1988). The social psychology of driving behavior: Communicative aspects of joint-action. *Mid-American Review of Sociology, 13,* 59-67.
5. Schelling, T. C. (1980). *The strategy of conflict*. Harvard University Press.
6. Brown, B. & Laurier, E. (2017). The trouble with autopilots: Assisted and autonomous driving on the social road. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 416-429.
7. Misyak, J., Noguchi, T., & Chater, N. (2016). Instantaneous conventions: The emergence of flexible communicative signals. *Psychological Science, 27,* 1550-1561.
8. Misyak, J. B., Melkonyan, T., Zeitoun, H., & Chater, N. (2014). Unwritten rules: Virtual bargaining underpins social interaction, culture, and society. *Trends in Cognitive Sciences, 18,* 512-519.
9. Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (Gold, N. & Sugden, R., eds), Princeton University Press.
10. Shafto, P., Goodman, N. D., & Griffiths, T. L. (2014). A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology*, *71*, 55-89.
11. Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818-829.
12. Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PloS One*, *8*(*11*), e79876.
13. Shashua, A. (2016). Sensing and beyond: Towards full autonomous driving. Powerpoint presentation, http://ir.mobileye.com/investor-relations/events-and-presentations/CES-2016-Presentation
14. Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, *352*(*6293*), 1573-1576.
15. Shariff, A., Bonnefon, J. F., & Rahwan, I. (2017). Psychological roadblocks to the adoption of self-driving vehicles. *Nature Human Behaviour*, *1*(10), 694.