Obava ze ztráty vlastního aktérství a absence morálního aktéra u autonomních dopravních prostředků[1]

Robin Kopecký

Vysoká škola uměleckoprůmyslová v Praze
Nám. J. Palacha 80, 116 93 Praha 1
Centrum Karla Čapka pro studium hodnot ve vědě a technice
https://www.cevast.org/
robin.kopecky@icloud.com

Michaela Jirout Košová

Centrum Karla Čapka pro studium hodnot ve vědě a technice
https://www.cevast.org/
michaela.kosa@natur.cuni.cz

Diskuse o filosofických problémech spojených s nástupem autonomní dopravy je často spojena s negativy a je vedena v duchu techno-pesimismu. Významný prostor v této debatě je věnován aplikaci vozíkového dilematu („*trolley problem*“) do řídících algoritmů vozů. V této eseji se však zamýšlíme nad dalším problémem našich morálních intuic – otázkou morálního aktérství. Pocit morální odpovědnosti si žádá faktickou kontrolu nad situací a kauzální odpovědnost. O tu bývalý řidič jako pasažér autonomního vozu přichází, ale přesto může cítit vinu. Prioritně bychom se měli soustřeďovat na morální psychologii uživatele, spíše než na hledání optima kolizí pro algoritmy. V eseji poukazujeme na historické paralely absence aktéra v automatických systémech i na to, že lidé snáze přijímají autonomní nákladní dopravu než dopravu osob. Dále nastiňujeme možné způsoby řešení přesunu odpovědnosti, a to nejprve prostřednictvím analogie s řidičem přicházejícím o aktérství nástupem do vozu hromadné dopravy. Možný operátor autonomních vozů, který na provoz dohlíží místo skutečného řidiče, může představovat dočasné opatření. Skutečnou alternativou morální odpovědnosti by se

1 Tento článek vznikl v rámci grantového projektu TAČR TL03000549 *Umění a design jako řešení změn v dopravě s nástupem autonomního řízení.*

ale mohl stát fiktivní morální aktér dosazený provozovatelem vozů. Oproštění se od vlastnictví autonomního vozu osvobozuje lidi od konečného vnímání morální odpovědnosti, takže krátkodobé výpůjčky, sdílení nebo hromadná doprava by mohly být další vhodnou odpovědí na tento psychologický fenomén.

Klíčová slova: aktérství, autonomní vozidla, morální odpovědnost, morální intuice

Uvedení technologických inovací v podobě propojeného a autonomního řízení pro individuální a hromadnou dopravu s sebou nese kromě jiného příležitosti pro pozitivní změny v podobě snížení dopadu dopravy na životní prostředí, zvýšení bezpečnosti pro cestující, a také finanční a časové úspory.[2] Přesto je společenská debata o těchto inovacích vedena v rovině možných negativ.[3]

Celé subžánry vědeckofantastické literatury se zabývají konfliktem mezi protagonisty a novými technologiemi. Konflikt představuje pole možností pro atraktivní děj, zatímco bezproblémové, čistě techno-optimistické prostředí je inertní a čtenářsky nezajímavé. Lze proto očekávat, že se i společenská diskuse bude ubírat techno-pesimistickým směrem, zaměří se na negativa a konkrétní etické otázky. Ačkoli v tomto textu řešíme namísto strojových zabijáků či všeobjímající virtuální reality ze science fiction obavy všedního dne, které odrážejí naše lidské morální intuice, sci-fi kultura hraje přímou kauzální roli i v této oblasti, jak dokládáme v následujícím odstavci.

Mohli bychom očekávat, že se čtenáři a fanoušci fantastiky, vzhledem k jejím odkazům na zneužití technologií, odlidštění, limitace a další problémy s fiktivní technologickou budoucností, stanou obezřetnějšími při morálním zvažování přínosů futuristických inovací. Na základě empirického šetření zkoumajícího morální schvalování technologického vylepšování člověka (enhancement) autoři Laakasuo a jiní zjistili, že jsou znalosti a příznivý vztah k science fiction s tímto souhlasem asociovány pozitivně, tedy že lidé s fantastikou lépe obeznámení morálně více schvalují enhancement. Otevřenost vůči zkušenosti (z pětifaktorového modelu osobnosti) s tímto schvalováním naopak asociována nebyla.[4]

2 Talebpour & Mahmassani (2016).
3 Kopecký (2019).
4 Laakasuo et al. (2018).

Přejděme ale od enhancementu a beletrie k hlavnímu tématu této eseje, kterým jsou překážky pro přijetí autonomní dopravy, a to nikoliv překážky technické, ve smyslu například optimalizace konstrukce vozovky a softwaru dopravních prostředků, ale překážky vyvstávající z naší lidské mysli a našich hodnot, a jejich možná řešení, která se pokusíme nastínit v závěru.

Tradiční a široce diskutovaný „vozíkový problém" (anglický termín „*trolley problem*" odkazuje na vůz tramvaje a překládá se jako vozíkové, tramvajové nebo drezínové dilema) představuje úkol pro programátory řídícího softwaru autonomních dopravních prostředků. Ti společně s filosofy hledají optimální morální rozhodovací pravidla, která by mohla být uplatněna v případě poruch nebo dopravních nehod, při nichž vznikají morální dilemata. Jde o situace kdy, pokud nezasáhneme, dojde jistě ke ztrátě na X životech. Pokud zasáhneme a riziko odvrátíme, následkem bude pravděpodobně nezamýšlená škoda na Y životech, přičemž X>Y.[5] V těchto dilematech stojí před morálním aktérem či strojem na jeho místě proti sobě omise (nejednat, a jistě nechat zemřít X lidí) a komise (jednat, a pravděpodobně nechat zemřít Y [X>Y] lidí). Jiným typem je „záchranářské dilema", ve kterém se má morální aktér nebo jeho případný robotický ekvivalent rozhodovat mezi dvěma komisemi: záchranou A, při které zemře B, anebo záchranou B, při které zemře A. V záchranářském dilematu nepředpokládáme omisi, při níž bychom nechali zemřít A i B.[6]

Myšlenkové experimenty zabývající se záchranou a nečinností vedoucí k úmrtí – vozíkový nebo záchranářský problém – však nejsou jedinými možnými typy problémů a domníváme se, že pro kontext společenského přijetí autonomního řízení nejsou ani těmi nejzávažnějšími. Úhelným kamenem zavádění nové technologie se může stát obyčejný uživatel automobilu, jehož ochota přijmout autonomní řízení může mít větší dopad, než nalezení optimálního řešení vozíkových problémů. Místo dilemat omisí a komisí pro aktéry, ať už lidské nebo jejich robotické alternativy, vyvstávají dva spojené problémy: *ztráta aktérství* a *absence morálního aktéra*. Tyto problémy jsou relevantní nejen pro tvůrce autonomních dopravních prostředků, ale i pro další filosofickou debatu týkající se

5 Foot (1967), Thomson (1985), Lin (2015), Černý (2020), Kopecky et al. (2022).

6 Sundvall et al. (2023).

s lidmi interagujících strojů s umělou inteligencí obecně. Návrhy řešení problémů s přijetím autonomní dopravy, které se podaří hypoteticky nalézt, tak budou následně aplikovatelné i v odlišných sférách etiky robotiky, například u pečovatelských nebo armádních robotů, které mají vyšší stupně volnosti.

Jak jsme už uvedli, za centrální problém nepovažujeme možné suboptimální nastavení morálního softwaru vozidla zodpovědného za řešení vozíkových dilemat z pohledu některých etických teorií (například konsekvencialismu nebo deontologie), ale řidičův strach z proměny jeho role aktivního aktéra na pouhý pasivní subjekt. Situaci navíc zhoršuje ztráta skutečného odpovědného morálního aktéra. Cítíme-li se totiž zodpovědní za nastalou situaci, chceme nad ní mít rovněž reálnou kontrolu, a přesune-li se zodpovědnost na někoho jiného, požadujeme, aby byl onou odpovědnou osobou skutečný morální aktér.

Na základě dat z Eurobarometru 2014 přinesli autoři Hudson, Orivska a Hunady pro tuto úvahu dva přínosné poznatky. Zaprvé zjistili, že je veřejnost otevřenější autonomním nákladním vozům než autonomním osobním vozům. To může být dáno tím, že se absencí aktéra i posádky jako takové v čistě nákladní dopravě situace morálního poměřování zjednoduší.[7] Dle druhého zjištění určuje technofobie k robotům měřená dle Nomura a jiných také negativní postoje k autonomním vozům.[8]

Jako tradiční historický příklad technofobie a strachu ze ztráty aktérství můžeme uvést výtah, který je svým způsobem obdobný problému chybějícího morálního aktéra u autonomního vozidla. Obecnou veřejnost znepokojovala absence operátora v nově zavedených automatických výtazích, byl tedy zaveden náhradní (lidský) operátor. Jeho zodpovědnost se sice vázala pouze ke stisknutí tlačítka podlaží, cestující se však díky jeho přítomnosti cítili příjemněji a služeb výtahu využívali mnohem ochotněji. Tato historická kuriozita ilustruje lidskou potřebu být schopen identifikovat konkrétního aktéra, který ponese plnou morální odpovědnost za možná řešení potenciálně nebezpečných situací. To může souviset s vnitřní intuicí, že pouze lidský aktér zvládne mít situaci náležitě pod kontrolou, klíčovou roli zde ale pravděpodobně

7 Hudson, Orivska & Hunady (2019).
8 Nomura et al (2006).

hraje možnost ukázat na někoho reálného jako na nositele morální odpovědnosti. Užití „výtahového řešení“ u autonomního vozu za použití emulace role řidiče řídícím softwarem by zřejmě neuspokojilo lidské obavy, protože by dostatečně nezohledňovalo lidskou intuici týkající se chybějícího lidského morálního aktéra. Inspiraci pro vhodné řešení, které popíšeme v následujících odstavcích, bychom mohli najít u hromadné dopravy.

Řidiče během ovládání vozidla činí morálními aktéry nutnost čelit mnoha dílčím morálním dilematům, od dodržení předepsané rychlosti, pravidel pro předjíždění a parkování, až po závažná provinění, jako je jízda pod vlivem drog. V případě autonomních vozů se však z původních řidičů stávají cestující – pasažéři.[9] Tím přestávají být schopni následovat či porušovat relevantní pravidla a tedy i čelit příslušným morálním dilematům. Chování autonomního vozu bude v souladu s platnými regulacemi, které tato dilemata dopředu vyloučí (jde o analogii k užití hypotetické „morální pilulky“ bránící konzumentovi po požití chovat se v rozporu s morálkou). K obdobné proměně dochází u řidičů osobní dopravy při nástupu do vozidla dopravy hromadné: i zde dobrovolně mění svůj status morálního aktéra na status pouhého pasažéra. Identifikujme dva možné důvody pro obvykle bezproblémové přijetí tohoto posunu.

Prvním důvodem usnadňujícím daný přechod je nezměněná fyzická přítomnost řidiče ve většině prostředků hromadné dopravy. Je tak zachován morální aktér, který ovládá vozidlo a nese zodpovědnost a morální vinu za případnou nebezpečnou situaci. Případný přenos přímo morálně zodpovědné osoby do autonomní osobní dopravy se ukazuje příliš složitým až nemožným. Lze si však představit živého člověka bez reálné zodpovědnosti dosazovaného na přechodnou dobu „za volant“, tj. do role morálního aktéra, jinak autonomně řízeného vozidla. Dočasně zde, stejně jako v případě výtahu, znamená, dokud si společnost nezvykne. Druhý důvod, pro užívání autonomních osobních vozů již plně aplikovatelný, je možnost nebýt vlastníkem/provozovatelem autonomního hromadného/osobního vozidla. Cestující v hromadné dopravě nenesou žádnou právní, morální ani psychologickou odpovědnost, což z nich dělá doslova morální „černé pasažéry“.

9 Kopecký (2019).

V případě vlastnictví autonomního vozu budou bývalí řidiči pravděpodobně stále cítit psychologickou odpovědnost za následky jednání řídícího softwaru, a to i navzdory tomu, že již nebudou fyzicky ovládat volant, pedály, ani páky. Požadované ztráty psychologické odpovědnosti můžeme dosáhnout náhradou osobního vlastnictví vozu krátkodobými pronájmy nebo sdílením u specializovaných společností. Přesun vlastnictví autonomních vozů z majetku uživatelů zmenší vnímanou morální odpovědnost za případné morálně relevantní pochybení vozu při řešení dopravních situací a případných nehodách.

Toto řešení zároveň odpovídá na otázku, zda by měl být řídící software regulován státem, výrobcem (automobilkou) či softwarovými vývojáři – psychologická vina i morální odpovědnost padne na vlastníka a provozovatele sdílených vozů. Marginalizujeme morální aktérství pasažéra, ale zachováváme morální odpovědnost provozovatele.

Různou míru morální odpovědnosti „řidiče", nebo lépe řečeno posádky autonomního vozu v rovině viny nebo zásluh, zkoumali McManus a Rutchick[10] na modelových situacích adaptovaného vozíkového problému převzatých ze studií Bonnefona a jiných[11]. Participanti hodnotili vinu či zásluhu posádky v závislosti na sobeckém/altruistickém rozhodnutí vozu při kolizní situaci. Ukázalo se, že nejvíce viny nebo zásluh přisuzují očekávatelně posádce – „řidiči" vozu v případě manuálního řízení vozidla, a naopak nejméně v případě, že byl výrobcem plně nastaven řídící software. Z příkladů různých mezistupňů intervencí posádky a přednastaveného softwaru od výrobce vyplývá, že redukce aktérství na straně „řidiče" (posádky) vede ke snížení vnímané viny či zásluh posádky za špatné či dobré následky řešení kolizních situací. I když je tento empirický výzkum založen na vozíkových problémech komise a omise, naznačuje, že zde existuje možnost omezením kompetencí pasažérů přenést morální vinu z posádky na výrobce.[12] V navazujícím výzkumu Benneta a jiných hodnotili participanti přisuzování viny „řidiči" dle různých stupňů automatizace. I když se s narůstajícím stupněm automatizace přisuzovaná vina „řidiče" redukovala směrem

10 McManus & Rutchick (2019).

11 Viz Bonnefon, J. F., Shariff, A. & Rahwan, I. (2016): „The social dilemma of autonomous vehicles." Science 352 (6293): 1573–1576. Dostupné z: https://doi.org/10.1126/science.aaf2654.

12 McManus & Rutchick (2019).

k dalším zúčastněným stranám, jako například k zmíněnému výrobci, přesto ve všech případech morální vina padala dle názoru participantů i na „řidiče“ – posádku autonomního vozu.[13]

Závěrem shrňme, že by problém laických morálních intuic týkajících se ztráty vlastního aktérství mohl nalézt částečné řešení ve změně vlastnických vztahů směrem k využívání vozu formou krátkodobých výpůjček a sdílení, nebo většinovým nahrazením soukromé dopravy dopravou hromadnou. Cílem je přesun břemena morální odpovědnosti z bývalých řidičů na nové majitele. Druhý problém, a sice absenci morálního aktéra, můžeme zkusit řešit analogicky k obsluze výtahů přechodnou přítomností lidských operátorů v autonomních vozech hromadné dopravy. Dalším krokem by mohlo být zavedení fiktivních morálních aktérů, jako je firma provozovatele/vlastníka, nebo do budoucna dokonce dostatečně sofistikovaných strojů. To by však vyžadovalo zahájení společenské a filosofické diskuse o umělých bytostech jako možné extenzi pro morální aktéry.

Obecné problémy, jako je technofobie vůči robotům, mohou být vyřešeny prostřednictvím fenoménu habituace. Vyrovnáme-li se s pravděpodobně roky trvající dobou habituace na různé roboty ve společnosti na základě pozitivních osobních zkušeností (jak můžeme vidět například v japonské kultuře), může se naše obava z absence morálního aktéra postupně transformovat do pozitivního vztahu k novému typu uměléhо aktéra. Ochota ke sdílení a přechodu k hromadné dopravě na jedné straně, a morální hodnocení chování robotů a autonomních vozidel na straně druhé, musejí být dále náležitě empiricky zkoumány, a to například formou hodnocení jednotlivých návrhů skrze dotazníková šetření. Jenom tak bude možné zhodnotit efektivitu našich návrhů a posunout je k možné implementaci.

Literatura

Bennett, J. M., Challinor, K. L., Modesto, O. & Prabhakharan, P. (2020): „Attribution of blame of crash causation across varying levels of vehicle automation.“ *Safety Science* 132: 104968. Dostupné z: https://doi.org/10.1016/j.ssci.2020.104968.

13 Bennet et al.(2020).

Černý, D. (2020): *The Principle of Double Effect: A History and Philosophical Defense*. Routledge.

Foot, P. (2002): „The Problem of Abortion and the Doctrine of the Double Effect." In *Virtues and Vices and Other Essays in Moral Philosophy*, P. Foot, Clarendon Press, Oxford, 2002, (Původně publikováno v Oxford Review, 5, 1967.) s. 19–33.

Hudson, J., Orviska, M. & Hunady, J. (2019): „People's attitudes to autonomous vehicles." *Transportation research part A: policy and practice* 121: 164–176. Dostupné z: https://doi.org/10.1016/j.tra.2018.08.018.

Kopecký, R. (2019): „Morální problémy autonomních vozidel." *Filosofický časopis* 67 (2): 267–268.

Kopecky, R., Jirout Košová, M., Novotný, D. D., Flegr, J. & Černý, D. (2022): „How virtue signalling makes us better: Moral preferences with respect to autonomous vehicle type choices." *AI & Society* 38: 937–946. Dostupné z: https://doi.org/10.1007/s00146-022-01461-8.

Laakasuo, M., Drosinou, M., Koverola, M., Kunnari, A., Halonen, J., Lehtonen, N. & Palomäki, J. (2018): „What makes people approve or condemn mind upload technology? Untangling the effects of sexual disgust, purity and science fiction familiarity." *Palgrave Communications* 4 (1): 1–14. Dostupné z: https://doi.org/10.1057/s41599-018-0124-6.

Lin, P. (2015): „Why Ethics Matters for Autonomous Cars." In *Autonomous Driving. Technical, Legal and Social Aspects*, eds. M. Maurer et al., Springer, Berlin, Heidelberg, s. 69–85. Dostupné z: https://doi.org/10.1007/978-3-662-48847-8.

McManus, R. M. & Rutchick, A. M. (2019): „Autonomous vehicles and the attribution of moral responsibility." *Social psychological and personality science* 10 (3): 345–352. Dostupné z: https://doi.org/10.1177/1948550618755875.

Nomura, T., Suzuki, T., Kanda, T. & Kato, K. (2006): „Measurement of

anxiety toward robots.“ In *ROMAN 2006 – The 15th IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, Hatfield, UK, s. 372–377. Dostupné z: https://doi.org/ 10.1109/ROMAN.2006.314462.

Sundvall, J., Drosinou, M., Hannikainen, I., Elovaara, K., Halonen, J., Herzon, V., Kopecký, R., Jirout Košová, M., Koverola, M., Kunnari, A., Perander, S., Saikkonen, T., Palomäki, J. & Laakasuo, M. (2023). „Innocence over utilitarianism: Heightened moral standards for robots in rescue dilemmas.“ *European Journal of Social Psychology* 53 (4): 779–804. Dostupné z: https://doi.org/10.1002/ejsp.2936.

Thomson, J. J. (1985): „The trolley problem.“ *The Yale Law Journal* 94: 1395–1415.

Talebpour, A. & Mahmassani, H. S. (2016): „Influence of connected and autonomous vehicles on traffic flow stability and throughput.“ *Transportation Research Part C: Emerging Technologies* 71: 143–163. Dostupné z: https://doi.org/10.1016/j.trc.2016.07.007.

Abstract

Fear of losing personal agency and the absence of a moral agent in autonomous transportation vehicles

Discussion about philosophical problems associated with the rise of autonomous driving is often associated with negatives and is conducted in the spirit of techno-pessimism. A significant space in this debate is dedicated to the application of the “trolley problem” to vehicle control algorithms. However, in this essay, we reflect on another problem of our moral intuitions – the question of moral agency. The feeling of moral responsibility demands factual control over the situation and causal responsibility. The former driver, now a passenger in an autonomous vehicle, loses this control, but may still experience guilt. Primarily, we should focus on the moral psychology of the user rather than searching for collision optima for algorithms. In the essay, we point out historical parallels of the absence of agency in automated systems and the fact that people are more accepting of autonomous freight transportation than personal transportation. Furthermore, we outline possible ways to shift responsibility, starting with an analogy to a driver losing agency when entering public transportation. A possible operator of autonomous vehicles, overseeing the operation instead of a real driver, could serve as a temporary solution. However, a fictional moral agent appointed by the vehicle operator could potentially become a genuine alternative to moral responsibili-

ty. Detachment from ownership of an autonomous vehicle frees individuals from a final perception of moral responsibility, so short-term rentals, sharing, or public transportation could be suitable responses to this psychological phenomenon.

Key words: agency, autonomous vehicles, moral responsibility, moral intuitions

Kopecký, R. & Jirout Košová, M. (2023): „Obava ze ztráty vlastního aktérství a absence morálního aktéra u autonomních dopravních prostředků.“ *Filosofie dnes* 15 (1): 52–61. Dostupné z www.filosofiednes.ff.uhk.cz