BS"D

The Trolley Problem Just Got Digital

Ethical Dilemmas in Programming Autonomous Vehicles

Rabbi Mois Navon[1]
Founding Engineer, Mobileye
Doctoral Student in Jewish Philosophy, Bar Ilan University
email: mois.navon@divreinavon.com
website: http://www.divreinavon.com

The following discussion is provided for educational purposes and is not meant to dictate how people or machines should drive.

---

[1] I would like to thank the rabbis at R. Asher Weiss' Institute for Research in Technological Innovation in Halacha (Jerusalem) for their insights and help in working out several issues mentioned in this article. I also thank R. Y. Medan of Yeshivat Har Etzion for his insights and discussions. I thank David Eisen (Beit Shemesh) for bringing to my attention many of the sources used in this article, as well as for the discussions on the issues herein. Finally, I thank my friends who read and commented on the draft version of this article: Maier Becker, Israel Belfer, David Guedalia, Michael Kara-Ivanov, Levi Kitrossky.

"*Great peace have they that love Your Torah; For them there is no stumbling*" (Ps. 119).

Abstract

Many a class on ethics opens with the renowned Trolley Problem. This ethical dilemma has been used to introduce the classical approaches of utilitarian versus deontological ethics. Now, though the Trolley Problem has had some real-life applications, the advent of the autonomous vehicle has just made the Trolley Problem very real. Autonomous vehicles will be programmed to make life-and-death decisions, and the burning question is how should they be programmed? This article seeks to provide a Jewish approach to this, and related, ethical dilemmas by employing classical Jewish sources (Mishna, Gemara, Bavli, Yerushalmi, Rishonim, Achronim, etc.) as well as modern rabbinic thought that considers the difference between a human driver and a digital one. The halachic analysis is accompanied by a discussion of the computing elements of the autonomous vehicle in an attempt to grapple with the intriguing, if not vexing, ethical dilemmas.

Bio

Rabbi Mois Navon is one of the founding engineers of Mobileye. Having designed the System-On-A-Chip and holding several patents in the hardware, he is intimately familiar with the system that is driving – literally – the autonomous car revolution. In addition, Mois is known as the "Rabbi of Mobileye," answering halachic questions on a personal and corporate level, giving a daily halacha and teaching a well-attended weekly shiur to religious and non-religious employees. This article is the result of one of those shiurim.

דילמת הקרון – מבחן המציאות
שאלות אתיות בתכנות הרכב האוטונומי

הרב מואיז נבון
מהנדס מייסד של מובילאיי
סטודנט לתואר שלשי בפילוסופיה יהודית, אוניברסיטת בבר אילן
מייל: mois.navon@divreinavon
אתר: http://www.divreinavon.com

תקציר

קורסים רבים באתיקה נפתחים בדילמת הקרון המפורסמת. דילמה אתית זו שימשה כדי להציג את הגישות הקלאסיות של האתיקה התועלתנית לעומת האתיקה הדאונטולוגית .עד כה, דילמת הקרון נותרה בעיקר אקדמית. אולם כעת, עם הופעתו של הרכב האוטונומי, דילמה זו נראית קרובה מתמיד למציאות. כלי רכב אוטונומיים יהיו מתוכנתים לבצע החלטות של חיים ומוות והשאלה הבוערת היא: איך הם צריכים להיות מתוכנתים? מאמר זה מבקש לספק גישה יהודית לדילמה זו, ולדילמות מוסריות הקשורות אליה, באמצעות שימוש במקורות יהודיים קלאסיים (משנה, גמרא, בבלי, ירושלמי, ראשונים אחרונים ועוד) וכן שימוש במקורות מחשבה רבנית מודרנית, אשר עושים אבחנה בין נהג אנושי לנהג דיגיטלי. הניתוח ההלכתי מלווה בדיון בנושא יסודות המחשוב של הרכב האוטונומי, בניסיון להתמודד עם הדילמות המוסריות המסקרנות, אם לא מביכות, אשר עולות.

ביו

הרב מואיז נבון הינו אחד מהמהנדסים המייסדים של מובילאיי. בתור מי שתכנן את השבב (SOC) והינו בעל מספר פטנטים בחומרה, הוא מעורה היטב במערכת המניעה את מהפכת הרכב האוטונומי. בנוסף, מואיז ידוע כ"רב של מובילאיי", הוא נותן מענה לשאלות הלכתיות ברמה האישית והארגונית, מעביר הלכה יומית ומלמד שיעור שבועי המורכב בעובדים דתיים וחילונים . מאמר זה הוא נכתב בעקבות אחד מהשיעורים האלה.

Introduction

The World Health Organization[2] estimates that approximately 1.25 million people die in traffic accidents around the world every year. They estimate that over 20 million people are injured in traffic accidents around the world every year. These numbers are about to practically disappear, for with the advent of the autonomous vehicle will come the removal of the number-one factor in car accidents: the human. This does not mean, however, that autonomous vehicles will not have to navigate ethically challenging situations: a truck may drop its payload into the road, a person may unexpectedly cross the street, etc., thus forcing the autonomous vehicle to have to decide who will live and who will die.[3] There are many such situations that we can conjure up, but perhaps the most famous is the one known as the Trolley Problem.

The Trolley Problem

> *You are driving a trolley that has lost its brakes. It is now hurtling down its track upon which five men are tied. While the trolley has no way to stop, you have the ability to throw a track switch that will divert the trolley from its current track to a parallel track, thus saving the five from certain death. On the parallel track, however, is tied a single man who will consequently be killed. What is the right thing for you to do?*

The above dilemma, originally formulated as an ethical thought experiment,[4] has ethicists divided into two camps: 1) those who look at the "utility" of the outcome – in this case, saving more people – and are thus known as utilitarians[5]; and 2) those who make their decision based on rules – in this case "thou shalt not murder" – and are thus known as deontologists (*deon* being Greek for duty).

---

[2] World Health Organization, Road Traffic Injuries Fact Sheet, Feb. 19, 2018. http://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[3] See P. Lin, "Why Ethics Matters for Autonomous Cars," *Autonomous Driving* (Springer, 2015), https://link.springer.com/content/pdf/10.1007%2F978-3-662-45854-9_4.pdf

[4] Philippa Foot, "The Problem of Abortion and the Doctrine of Double Effect," *Oxford Review* 5 (1967), p. 8.

[5] For the sake completeness, utilitarianism, which seeks the action that will result in greatest happiness for the greatest number of people, is the most popular form of a more general approach known as consequentialism (See M. J. Harris, "Consequentialism, Deontologism, and the Case of Sheva ben Bikhri," *The Torah u-Madda Journal* [15/2008-09], p. 68).

While the utilitarian approach is appealing, since saving as many people as possible always seems like a good thing, we enter murky territory when we begin to attach names or titles to the people on the track. For example, what if the single man is a head of state? Should he take precedence over five ordinary citizens?[6] Before tackling this problem, let us look at a simpler version of the problem known as the Tunnel Problem.[7] Here, a driver is approaching a single-lane tunnel, and there is a pedestrian in the road. The driver does not have time to brake and is left only with the choice of running over the pedestrian or killing himself by driving into a wall.

One Against One

In pitting one individual against another, the quantitative element is removed from the equation, thus allowing us to focus on the more salient aspects of the ethical dilemma. The Mishna (Ohalot 7:6), addressing the one-against-one dilemma, teaches that man does not have the wherewithal to judge between individuals, and so "one life is not set aside for another." Now, while most people are quite comfortable with this egalitarian stance, they get apprehensive when, as in the Tunnel Problem, the question gets personal. That is, if the choice is between running over a stranger or sacrificing your own life, what do you do?

The Talmud (Pes. 25b)[8] formulates this dilemma as follows: "The governor of a city said, 'go and kill Ploni or you will be killed.'" What do you do? Rabbah responds that one must give his own life rather than commit murder, for "In what way do you see that your blood is redder than his? Perhaps his blood is redder?"[9] Egalitarianism, the Talmud establishes, applies even when it gets personal.[10]

---

[6] For more examples, see: http://moralmachine.mit.edu/.

[7] Jason Millar, "You Should Have A Say In Your Robot Car's Code Of Ethics," *Wired*, Sep. 2, 2014, https://www.wired.com/2014/09/set-the-ethics-robot-car/.

[8] Similarly, San. 74b.

[9] "Redder blood" means that one is more valuable to the world in his capacity to do more in the way of fulfilling the will of the Creator (Hidushei Talmidei Rabbeinu, Avoda Zara 28a; Sefer HaHinuch #296).

[10] Worthy of mention is that, while there are sources that do prioritize between various people in society (Horayot 3:6-8), these are inapplicable when adjudging who will die, as opposed to who will be saved (see R. Kook's explanation in fn. 39). Regarding the inapplicability of these prioritizations in programming an autonomous vehicle, see R. Y. M. Malka, *Halachic, Ethical and Governmental Challenges in the Development of the*

That being said, the Tosafot (San. 74b, *ve'ha*) note that while self-sacrifice is indeed demanded when one will *actively* murder another, passive action is different. They hold that if the governor said, "allow me to throw you onto a baby such that you will end up crushing him to death," one would not be demanded to sacrifice one's own life, "for he did no action."

R. Haim Soloveitchik[11] explains the Tosafot's reasoning as follows: since both people are equal, neither one has the right to kill the other, but by the same token, neither one is obligated to give his own life to save the other. That is, as long as one does not actively kill another, he may remain passive – "*shev v'al taaseh*."[12] R. Haim goes on to explain that, while there is logic to the Tosafot's argument, the Rambam uses the very same principle – that all people are equal (*shekulin hen*) – to argue that even passive killing is not permissible. That is, precisely because all people are equal, there is no justification to set aside one life for another (*ain ba din dehiyah*). Consequently, according to the Rambam there is no difference between actively killing or passively killing – in all such cases, one must sacrifice oneself.

This brings us to another Talmudic scenario that pits one individual's life against another.

> Two people were walking [far from civilization] and only one of them has a canteen of water. If both drink, they will [both] die, but if only one drinks, he can reach civilization. Ben Petora taught: It is better that both should drink and die, rather than that one should behold his companion's death. [And so it was] until R. Akiva came and taught: 'that your brother may live with you' (Lev. 25:36) [means] your life takes precedence over his life.[13]

Autonomous Vehicle (Jerusalem: The Institute for Research in Technological Innovation in Halacha, Kislev 5778 [Hebrew]), p. 34, and Sec. "Priorities in Saving Lives", pp. 195-199. See also Tzitz Eliezer (17:72:15).

[11] Hidushei R. Haim Halevi, Yesodei HaTorah, Ch. 5.

[12] See further (Sec. "The Altruist") on the permissibility of voluntarily sacrificing oneself for others.

[13] Baba Metzia 62a. For further discussion on the positions of Ben Petora and R. Akiva, see Hatam Sofer (Baba Metzia 62a), Iggrot Moshe (YD I:145), Tzitz Eliezer (17:72), R. Medan ("On a Canteen of Water", Daf Kesher 766 - http://gush.net/dk//1to899/766mamar.htm). As an important aside, R. Medan notes that, while the normative halacha accords with R. Akiva, there are cases – e.g., leaders, soldiers – where one must act according to Ben Petora.

The Minhat Hinuch writes that R. Akiva's position reflects the conclusion of the Tosafot, namely that "one is not required to save his friend at the expense of his own life."[14] Now, although it is true that both the Tosafot and R. Akiva allow for an individual to refrain from sacrificing his own life while the other dies, R. Haim notes an important difference.[15] In the case of the Tosafot, the individual, though passive, is nevertheless very much involved in a murder (*retzicha*), whereas in the case of R. Akiva, the individual simply declines the option of saving (*hatzala*) the other.[16] Furthermore, explains R. Haim, the fact that R. Akiva needed to bring a scriptural verse to permit not saving a life, an act that in no way falls under the legal rubric of murder, demonstrates that in the case where there is an act of murder, albeit passive, one cannot rely on anything less than scriptural authority. R. Haim concludes, "In any case where murder is involved, even if entirely passive, one must sacrifice oneself (*yehareg v'al ya'avor*)."

Having said this, there is a dispute regarding whether the imperative to sacrifice oneself (*yehareg v'al ya'avor*) demands only that one *passively* submit to death or if one must *actively* take one's own life (i.e., suicide) in order to avoid the sinful act (e.g., murder).[17] While some authorities forbid one to commit suicide rather than committing a sin demanding self-sacrifice and some authorities permit one to do so, the Ramban champions the position that one must commit suicide.[18] Nevertheless, explains R. S. Dichovsky, in a case where one would be killing another passively, the killer is not considered an actual murderer, and thus even the Ramban would not demand he take his own life.[19] Accordingly, in the case of passively killing another, even the Rambam, who holds that one must give his life

[14] Minhat Hinuch, 295-296, #1.

[15] Hidushei R. Haim Halevi, *loc cit.*

[16] Similarly, Hatam Sofer (Baba Metzia 62a).

[17] See Encyclopedia Talmudit, Entry: *Yehareg V'al Ya'avor*, n. 119-121. See also R. M. Adler, "Definitions of Suicide," *Yeshurun* 13 (Hebrew).

[18] This is according to R. Elchanan Bonem Wasserman (Kovetz HeArot, Yevamot 48, esp. #5); however, Zichron Shmuel (65:32) holds that even the Ramban only *permits* suicide but does not obligate it.

[19] R. S. Dichovsky, "Precedence and Priority in Saving Lives according to Halacha," *Dinei Yisrael* 7 (Hebrew), p. 47. He brings R. Haim who states that one who kills another passively (e.g., thrown on a baby) is not considered a murderer.

passively (*yehareg*), would agree that there is no demand for one to *actively* commit suicide in order to avoid passively killing another.[20]

With these sources in mind, we can now return to the Tunnel Problem which has two sub-cases to be considered: 1) passive and 2) active.

1) If the street is perfectly straight and the driver is simply holding the steering wheel straight, this would be considered passive killing and the driver would not be obligated to actively take his own life.
2) If the street is curved such that the driver must actively turn the wheel into the curve, this would be considered active killing and the driver would have to give his own life (passively driving the car straight) to avoid running over the pedestrian.

Despite the justness of the above rulings, there is an important mitigating factor to be noted: the legality of the pedestrian. In *Halachic, Ethical and Governmental Challenges in the Development of the Autonomous Vehicle*, a book soon to be published under the auspices of R. Asher Weiss' Research Institute, the authors note,

> When an individual is driving according to the law and a pedestrian is illegally in the road, then the pedestrian's very presence, which would obligate the driver to sacrifice himself by driving off a cliff, defines the pedestrian as a "*rodef*" (a person pursuing another to kill him) … and consequently, the pedestrian has himself forfeited his own right to life. … [Similarly] is the case when a pedestrian bursts into the middle of the street.[21]

That is, if the pedestrian is not legally permitted to be in the street for whatever reason, he has no right to cause another person (i.e., the driver) to sacrifice his life; and thus, all would agree that the driver need not sacrifice his own life.[22]

---

[20] It is important to note here that, though the Rambam demands one be killed rather that passively killing another, this is because there is no way to determine whose blood is redder and not because there is no difference between passive and active killing of another. Indeed, as R. Haim noted (see previous footnote), the passive killer is not considered an actual murderer.

[21] Malka, Sec. "Implications of Din Rodef in Such Cases," pp. 175-177.

[22] One cannot help but be reminded of the first fatal pedestrian accident caused by an autonomous vehicle in which a woman attempted to cross a street where there was no crosswalk (see:

One Against Many

With the understandings gleaned from the one-against-one Tunnel Problem, let us now approach the one-against-many Trolley Problem. (Note that the Trolley Problem differs from the Tunnel Problem not only in that it pits one against many, but also in that the driver does not have the option of self-sacrifice). The primary source for this discussion is the Jerusalem Talmud (Terumot 8:4),[23] which we will refer to as the Marauders Case:

> A group of people were traveling and marauders chanced upon them saying, 'Hand over one of your group or we will kill you all.' Even if all will be killed, they may not hand over one soul. If they [i.e., the marauders] specified (*yichduhu*) an individual like Sheva Ben Bichri[24] then they [i.e., the group] hand him over and are not killed. Resh Lakish said [that this permit applies only if the specified individual] is liable for the death penalty like Sheva Ben Bichri. R. Yohanan [disagreed and] said that [the permit is applied] even if [the specified individual] is not liable for death as Sheva Ben Bichri.[25]

Setting aside for the moment the special case of the specified individual, this source unequivocally rejects utilitarianism, leading to uncomfortable implications when applied to

---

https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe). While there has been much discussion over whether the accident could have been avoided given the proper technology or configuration thereof, the moral argument here is that if the driver – human or otherwise – did identify the pedestrian but could only choose between running her over or killing the driver, the choice, in this case, would be to run over the pedestrian.

[23] Also, Tosefta Terumot 7:27.

[24] See Samuel II, Ch. 20.

[25] The Rambam (Hil. Yesodei HaTorah 5:5) holds like Resh Lakish and the Meiri (Beit Habehira, San. 72b) holds like R. Yohanan. For the sake of completeness, it is important to note that the Gemara concludes this discussion with a story of R. Yehoshua Ben Levi who convinced Ulla Bar Koshav to give himself over to marauders as per this ruling of "*yichduhu*" for which Eliyahu demonstrably condemned. As a result of this conclusion, even though the law permits handing over an individual in the special case of *yichduhu*, the Rambam writes that we "don't tell people to do so" and similarly the Meiri writes that it is "the way of the righteous is to delay" such an act. For an analysis of the Rambam's position see, R. Asher Weiss (Minhat Asher, Pes. #28). For a thorough analysis of this *sugya* see M. J. Harris, "Consequentialism, Deontologism, and the Case of Sheva ben Bikhri," *The Torah u-Madda Journal* (15/2008-09).

a driver on the road confronted with the Trolley Problem. In an attempt to deal with this dilemma, the Hazon Ish (San. #25), using what we will refer to as the Missile Case,[26] proposes that the utilitarian approach could possibly be applied if we could frame the dilemma as saving people as opposed to killing people. Due to the critical importance of his words, I bring them in *toto*, with my clarifications interspersed.

> Someone sees a missile heading toward a multitude of people and he can divert it to a different side such that it will kill only one person while on the other side the multitude will be saved. Conversely, if he does nothing, the multitude will be killed and the individual will remain alive. It is possible that this case is unlike the [Marauders] case wherein one person is handed over (*moser*) to be killed; for there the act of handing over someone is a brutal act of killing and the salvation of the others is not an inherent [lit.: natural] part of this act but is rather the indirect cause of their salvation.
>
> > [MN: That is, when the group hands over an individual to be killed, the group is saved indirectly, as the threat of the marauders is now removed. Consequently, the act of handing over the individual can be viewed in isolation as "a brutal act of killing." This is different than diverting the missile, in which case the very act of diverting the missile inherently saves the people under fire.]
>
> Also, [another important distinction to be made here is that, while the act of handing over the individual can be viewed in isolation], the salvation of the group is, in fact, directly connected to handing over the individual to be killed.
>
> > [MN: That is, the salvation of the group is dependent, inextricably, on the killing of the individual. Consequently, the act is more about killing than about saving. This stands in contradistinction from diverting the missile, an act which saves the multitude without – necessarily – killing anyone. That is, the act *qua*

[26] R. Dr. Michael Avraham, ("Separating Siamese Twins", *Asia*, 12, 3-4, Kislev 5750 [Hebrew], p. 184), notes that the Missile Case was conjured to discuss a car caught in a Trolley Problem. Accordingly, it assumes, like the Trolley Problem, that there is no option for self-sacrifice nor is there an option for the agent to walk away. (N. Rakover describes the Hazon Ish's motivation to be a case of self-sacrifice, though the Missile Case itself does not entertain that possibility – see *Self-Sacrifice – Sacrificing the Individual to Save the Many* (Jerusalem: The Library of Jewish Law, 2000) [Hebrew], p. 65).

act is not one of killing. It just so happens that – in this particular instance – the diverted missile will end up falling on an individual resulting in his death. To make this distinction clear, consider the following *Transplant* dilemma.[27] A doctor has five patients in need of various organs. If no organs are forthcoming, the five patients will all die. One day a healthy young man comes into the doctor's office for an annual check-up. The doctor can save the five by killing the one. The salvation of the many is directly dependent on, and inherently a part of, killing the individual who is a means to an end.[28] This is not the case when diverting the missile.]

Diverting the missile from one side to another is, in its essence, an act of salvation not connected in any way to killing the individual on the other side. It is merely incidental that there happens to be an individual there. Now, given that on one side the multitude will be killed and on the other side one will be killed, it is possible that we should strive to minimize losses as much as possible. Indeed, did not Lulianus and Papus who, [upon hearing of the blood libel in which the king threatened to kill all the Jews if someone did not take responsibility for the murder of his daughter, sacrificed themselves] to save Israel [i.e., minimize losses], as noted by Rashi (Taanit 18b, s.v., *BeLudkia*), and of whom it is said [in great praise] that no creature can stand in their holy place in Heaven.[29]

[MN: Here the notion of utility, or minimizing losses, is introduced as an additional consideration. That is, since we have a case where we are not directly committing a brutal act of killing, but rather saving people and incidentally an individual will die (i.e., there is, apparently, no violation of deontological considerations), perhaps, offers the Hazon Ish, we should include the utilitarian consideration (as learned from Lulianus and Papus). So while there is clearly value in the utilitarian consideration, it comes into play only after deontological considerations are maintained.]

---

[27] In Judith Jarvis Thomson, "The Trolley Problem," *Yale Law Journal* 94, 1985, p. 1399.

[28] Thomson (p. 1401) brings Kant to highlight the violation here: "Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only."

[29] See also Baba Batra 10b.

> However, here [in the Missile Case] it is weaker [than we have made it out to be], for one is killing with one's own hands (*hariga beyadayim*). And we have found that only handing over an individual is allowed [in the case when the marauders specify the individual to be handed over or, according to the stricter opinion, that the individual be liable for the death penalty] as was Sheva Ben Bichri.[30] And this needs investigation.
>
> > [MN: That is, up until now, it has been argued that diverting the missile is an act of salvation carrying with it the incidental killing of an individual. Here, however, the Hazon Ish expresses reservations since diverting the missile is nonetheless "killing with one's own hands," for one does, at the end of the day, point the missile at the individual. The Hazon Ish then explains that only in the special case where the individual is liable for the death penalty can he be handed over to marauders to save the many. Accordingly, one cannot divert the missile that will kill an individual and save the many, given that the individual is not liable for the death penalty.]

The Hazon Ish himself, then, remains undecided over the possibility of reframing the Trolley Problem in terms of saving versus killing, concluding that the issue needs investigation.

R. Asher Weiss (Minhat Asher, Pes. #28, 8) responds to the call for such investigation, noting that, indeed, it is hard to understand the distinction between a "brutal act" versus a "saving act," for in the final analysis, "when one kills a person by his own force (*kocho*) it is a violation of murder." R. Weiss then offers three possible ways to understand the Hazon Ish's proposal (paraphrased here):

1) If one were to abstain from diverting the missile away from the multitude, one could be guilty of violating two biblical commandments – "do not stand idly by your neighbor's blood" and "return [his health] to him" – which demand that one help his fellow man. Perhaps the Hazon Ish is arguing that being in violation of these two commandments, for each and every one of the people in the multitude,

[30] See R. Asher Weiss (Minhat Asher, Pes. #28, Seif 3-7, esp. 6) who explains that only because the individual is adjudged a *rodef* whose blood is less red can we hand him over.

outweighs being in violation of actively (*kum v'aseh*) killing a single individual. "And this still needs investigation."

2) We might look at the Missile Case as one in which all of the people, including the individual, would be killed. While this scenario is comparable to the marauders demanding one individual lest they kill the whole group, in which case one is not allowed to hand over an innocent person, one might argue that diverting the missile is an act of saving as opposed to handing over an individual which is an act of killing. [And indeed, in this case where all would be killed, we are certainly obligated to minimize deaths].[31] But it is clear that the Hazon Ish was not talking about such a scenario. "And this needs a lot of investigation."
3) It seems we can make sense of the Hazon Ish's proposal to divert the missile if we assume it is diverted passively. That is, one may not actively bend the path of the missile to kill the individual as this would be considered one's own force (*kocho*), which is a clear violation of killing with one's own hands (*hariga beyadayim*) and prohibited even to save the many. However, one could conceivably place a "shield" over the multitude such that the missile would bounce off of it and land on the individual. Placing a shield would only indirectly cause the death of the individual and thus would not be considered one's own force (*kocho*).[32] "And all of this needs a lot of investigation."

Though R. Weiss considers three possible ways to affirm the Hazon Ish's reframing of the Trolley Problem as an act of saving versus one of killing, ultimately he remains, like the Hazon Ish, inconclusive.

In stark contrast to R. Weiss, the Tzitz Eliezer (15:70) also answers the Hazon Ish's call for investigation and, referring explicitly to a car driver caught in the Trolley Problem,[33] comes

[31] See fn. 39. See Malka, p. 154, fn. 177.

[32] Note that the point here is not to work out the technical implementation of such a scenario, but to understand the parameters of moral actions – i.e., saving the many through the passive killing of the few.

[33] It should be noted that, though the Tzitz Eliezer speaks of a car in which the driver slams on the brakes to avoid the many and then reverses only to kill an individual behind him, this is clearly not the scenario that he intends; for obviously if the brakes had been applied to the point that one could reverse, there is no dilemma here. Rather he is, as he states explicitly, applying the Hazon Ish's Missile Case to the real-life case of a car that must choose between the one and the many. All that can be said on this is that apparently the Tzitz Eliezer didn't drive.

to very decisive conclusions. The Tzitz Eliezer begins by quoting the Marauders Case to show that one cannot save the many at the expense of the individual. He then brings the words of Rabbeinu Yona[34] who explains the reasoning behind the "whose blood is redder" argument. According to Rabbeinu Yona, you might argue that since the "redness"[35] of your neighbor's blood stands in question compared to your own, hence your neighbor should be killed and not you. Nevertheless, you must remain passive (*shev v'al taaseh*), for "one is to refrain from doing any sin with one's own hands." On this the Tzitz Eliezer writes,

> Rabbeinu Yona has laid down for us a critical general principle [to illuminate] the guiding principle that our sages have set for us in relating to questions of life and death, and that is to choose to be in a state of passivity (*shev v'al taaseh*). This is true regardless of which side of a scenario we find ourselves; one must remain passive when it is impossible to resolve whose blood is redder, being ever guided by the principle to refrain from doing any sin with our own hands.

With this, the Tzitz Eliezer goes on the attack against the Hazon Ish, explaining that this guiding principle to refrain from sinning with our own hands applies equally whether the scenario is one-against-one or one-against-many. In all cases, one must remain passive (*shev v'al taaseh*). He notes that the Missile scenario is precisely the same as that of a car driving toward a group of people where the driver could change directions to hit an unrelated individual and thus save the multitude. Performing any action that would result in actively (*kum v'aseh*) killing someone is impermissible, regardless of the intent to save the many. Consequently, explains the Tzitz Eliezer, "in the case of the Hazon Ish we must resolutely decide (*lifsok behehletiut*) to remain passive and not actively divert the missile."

The Tzitz Eliezer leaves us with the following decisive, deontological, yet difficult words: "In any case of certain killing, there is no distinction between the individual and the multitude, for we do not say that the multitude is to be favored." On the one hand, we can take comfort in operating according to clear-cut rules. On the other hand, imagining ten, five, or even two people killed because one may not turn the steering wheel is a hard pill to swallow. Perhaps it helps to know that what is being articulated here is nothing less than

[34] Hidushei Talmidei Rabbeinu, Avoda Zara 28a.

[35] See fn. 9.

one of the great ideals the Bible has bequeathed to the Western world: the inestimable value of the individual.

Across the pagan world the individual was sacrificed, literally, for the sake of the many. In contrast, writes R. Moshe Avigdor Amiel in his *Ethics and Legality in Jewish Law*, the Bible opens with the revolutionary notion that man was created in the image of God, and just as God is singular and unique, so too is man singular and unique.[36] The Talmud (San. 4:5) phrases it thus: Man was created alone to teach you that whoever destroys a single soul is as if he had destroyed a whole world, and whoever saves a single soul is as if he had saved a whole world.[37] Applying this to the Trolley Problem, Rav Kook explains that "It is beyond our power to estimate the value of the *whole world* that is the individual versus the *whole world* that is the multitude."[38] That is, if every person is of infinite value, then pitting one infinity against many infinities still results in an identity that does not allow us to favor one over the other in taking a life.[39]

Utility

Even while accepting the inestimable value of the individual, there is a voice that says careening into a mass of people to refrain from turning the steering wheel into an individual

[36] See R. Y. Beasely, "The Importance of One," http://etzion.org.il/en/importance-one.

[37] That the correct version is "whoever saves a soul" and not "whoever saves a soul in Israel," see Ephraim E. Urbach. "'Kol Ha-Meqayyem Nefesh Ahat...' — Development of the Version, Vicissitudes of Censorship, and Business Manipulations of Printers", *Tarbiẓ* no.3, 1971, pp. 268–284 [Hebrew]. *JSTOR*, www.jstor.org/stable/23593213.

[38] Responsa Mishpatei Cohen #143.

[39] Please note that this does not mean that Judaism does not recognize the value in saving the many, but rather, as mentioned earlier: while there is clearly a value in the utilitarian consideration, it comes in to play only after deontological considerations are maintained. So, for example, if the choice was not about killing but about saving, such that one could make only a single boat trip to save either one person on Island A or ten people on Island B, clearly saving the many would triumph. Similarly, the Tzitz Eliezer (20:2) himself – in a case of quadruplets where one of the four fetuses must be aborted to save the rest – explicitly rejects the applying the marauder case since it is not considered murder to kill the fetus. R. Kook (Mishpatei Cohen #144) distinguishes between lifesaving cases where we can apply "common-sense" (*umdena*) versus capital cases (*dinei nefashot*) where we cannot. See discussion thereon in Rakover, *Self-Sacrifice*, p. 51-58, esp. p. 58.

just doesn't make sense.[40] This voice is articulated by R. Yaakov Medan who takes what he calls a common-sense approach to the problem. He asks rhetorically who would have an easier time explaining their actions to an impartial judge investigating such a car accident: the driver who ran over the one person or the driver who ran over the many? The common-sense approach to ethical problems finds support in the Responsa of the Radvaz (3:627) which cites the biblical verse, "All its paths are paths of peace," and explains that "the ordinances of our Torah must accord with intellect and reason." This, of course, does not mean that mere "reason" or "common-sense" can determine Jewish ethics but simply that the norms determined by the Torah are to accord with reason.[41] Consequently, R. Medan must explain how his common-sense approach can be reconciled with the halachic sources used in this discussion.

The primary source asserting that the many cannot be saved at the expense of the one is, as we have seen, the Marauders Case.[42] The Hazon Ish, the Tzitz Eliezer and Rav Asher Weiss all employ this case as the starting point for their discussion on this issue.[43]

---

[40] This appears to be the intuitive response of most people (See Skulmowski, et. al, "Forced-choice decision-making in modified Trolley Problem situations," *Frontiers in Behavioral Neuroscience* Vol. 8, 2014. https://www.frontiersin.org/article/10.3389/fnbeh.2014.00426). Important to note here is that, while a driver doing nothing can be considered "passively" killing, such a decision is, nevertheless, a more involved action than "letting die," as, for example, when a bystander can switch a trolley from its current track or walk away "and let die" the many (Thomson, pp. 1396-8).

[41] Similarly, secular ethicists explain that, "one cannot infer an ethical theory from people's intuitions … Normative conclusions must be supplied by ethical theories. The empirical investigation only yields which of these theories is more aligned with society's practices and people's intuitions …" (Bergmann, et al, "Autonomous Vehicles Require Socio-Political Acceptance—An Empirical and Philosophical Perspective on the Problem of Moral Decision Making," *Frontiers in Behavioral Neuroscience* Vol. 12, 2018. https://www.frontiersin.org/article/10.3389/fnbeh.2018.00031).

[42] For example, see Responsa Bach (43), Responsa Iggrot Moshe (YD II:60) who employ the Marauders Case. See also Responsa Seridei Eish (2:38) and Rakover, *Self-Sacrifice*, Ch. 2.

[43] So too R. Bleich, "Sacrificing the Few to Save the Many," *Tradition* 43:1 (Spring 2010). As an important side note: R. Bleich discusses the permissibility of shooting down the plane being used to hit the Pentagon in the September 11, 2001 attack. He concludes that one cannot kill the few to save the many and thus one could not shoot down the plane. While he is on strong footing to prohibit sacrificing the few in favor of the many, I would humbly disagree to its application in the September 11 case. There the issue is not about saving the many, it is about saving the country. The airplanes were simply missiles that happened to have innocent people in them. A missile attack is an attack on the very sovereignty of the nation and must be dealt with accordingly. When seeking to preserve the integrity and welfare of the nation, different rules apply than when

Furthermore, all concur that, barring some mitigating circumstance (e.g., "inherent saving"), the Marauders Case teaches a hard and fast rule that the many cannot be saved at the expense of the one. R. Medan, expressing clear discomfort in going against such giants in halacha, nevertheless says, "though I am merely dust at their feet, I humbly beg to differ."[44] Referring to his article, "Mishnat Hassidim," he explains that the Marauders Case is brought only in the context of war or persecution.[45] It is a teaching to be employed in times of national adversity to maintain the unity, integrity and spirit of the people. It does not, explains R. Medan, apply to peacetime deliberations.[46]

This then brings us to the "whose blood is redder" argument. As we have seen, one life cannot be saved at the expense of another, for one knows not "whose blood is redder." Furthermore, just as one cannot determine the value of one individual versus that of another individual, "by the same reasoning," writes the Tzitz Eliezer (15:70), neither can one determine the value of one individual versus that of the multitude.[47] In the words of Rav Kook, "We cannot evaluate one soul against other souls, even though they be many, for we have not the authority to determine who shall die (*umdena be'dehiyat nefashot*)."[48] As of this writing, R. Medan has yet to provide support for his claim to dispose of the "whose blood is redder" argument.

---

dealing with civil issues (e.g., the Trolley Problem). For a discussion on this topic see: Y. Hazony, *The Dawn* (Jerusalem: Shalem Press, 2000), Chs. 20-21. See also R. M. A. Amiel, "Nevuchei HaTekufa", #66 quoted in R. Y. Zisberg, "Saving Lives when Settling the Land of Israel", No. 150, 5763, pp. 236-7.

[44] Personal conversation.

[45] Alon Shevut 150, 5758,
http://asif.co.il/download/kitvey-et/alon%20shevut/alonshevut150/150chasidim.html

[46] R. Medan anticipates a claim against his understanding in that the same sources that bring the Marauders Case (Jerusalem Talmud Terumot 8:4; Rambam, Hil. Yesodei HaTorah 5:5), also bring a Molesters Case: Molesters come upon a group of women and say give us one woman to violate or we'll violate everyone: all must be violated, and not one given over. Here too is a case against sacrificing the one to save the many which appears not to be confined to wartime or persecution. Nevertheless, argues R. Medan, this law too was given to be applied only when the unity and integrity of the people is at stake. See also Harris, pp. 81-83, for a discussion on reading these sources as applying only when the community is threatened.

[47] See also Rakover, *Self-Sacrifice*, Ch. 4, esp. pp. 50-51.

[48] See fn. 38. See also Harris, p. 74.

R. Medan's common-sense approach notwithstanding, one could do worse than to offer the idea of *aveira lishma* – performing a sin for a greater good – to argue the utilitarian case.[49] That is, halacha does recognize that there are situations – very limited in scope – in which the ends justify the means. On the verse "In all your ways know Him," the Talmud (Ber. 63a) teaches "even in a sin."[50] Rashi explains this to mean that one can commit a sin if the action is for the sake of mitzvah, "like Eliyahu on Mt. Carmel [who brought sacrifices outside the Temple precincts in order to bring back the people to righteousness]."[51] In another place the Talmud (Hor. 10b) relates how Yael seduced Sisera, the enemy general, only to drive a stake through his head when their encounter was over. Being a married woman, Yael violated the prohibition of adultery, yet in committing her sin for a noble reason (*aveira lishma*), she is said to be more blessed than the matriarchs.[52]

Though these sources do appear to condone performing a transgression for the sake of the greater good, the commentators grapple over the conditions of its application. The Netziv explains that *aveira lishma* applies to all types of transgressions; however, to justify the act one must fulfill three conditions: a) attain no personal enjoyment/benefit from the act,[53] b) have the act thrust upon oneself, there being no good alternative,[54] and c) be steeped in Torah thought to be able to determine the worthiness of such an act.[55] R. Haim of Volozhin takes this last condition a step further stating that only a prophet can institute, temporarily,

---

[49] I offer this idea for, after analyzing the various options, the Trolley Problem leaves us in a difficult predicament. Thomson, too, reached this quandary and noted that while, indeed, sacrificing the individual is a "wrong" (p. 1405), nevertheless, it is "a wrong it is permissible for him to do." (p. 1412). That is, she is not willing to say that sacrificing the few to save the many is demanded, but she is also not willing to say that it is prohibited. Interestingly, such a gray resolution does exist within the context of halachic discourse: the *aveira lishma*.

[50] Berachot 63a.

[51] Maharitz Hiyut (Ber. 63a) brings an expanded text of Rashi to say that one should have the intent of fulfilling a mitzvah and thus, echoing the conclusion of the verse in Proverbs (3:6), "God will keep him on the straight path." This does not imply that Rashi gives *carte blanche* to such actions but may permit them as a temporary injunction (*horaat shaa*) of a prophet (Hevruta, Berachot 63a, n. 18).

[52] Also Nazir 23b. As to why Yael is greater, see Rashi (Nazir 23b), Meshech Hochmah (Ex. 17:8).

[53] Harhev Davar (Gen. 27:9, Num. 21:30) notes that punishment for the performance of the sin accrues only if the act is done with intent of personal enjoyment (i.e., not altruistically). Many align with this demand of altruism (see Hevruta, Horayot 10b, n.7).

[54] HaAmek Davar (Num. 15:41).

[55] HaAmek Davar (Deut. 4:3) explains that without being steeped in Torah one can easily end up unjustifiably sinning. Similarly, Tifferet Yisrael (Avot 2:1) notes the importance of this third condition.

an *aveira lishma*.[56] The Noda Beyehuda[57] establishes even more limitations, writing that such an act is permitted only if "all of Israel, from Hodu until Cush," will be saved, and even then, the act must have the sanction of the high court "and perhaps revelation as well." Furthermore, in contradistinction to others, he does not allow the application of *aviera lishma* when the sin is idol worship or spilling blood (e.g., as in driving over an individual to save the many).

Now, even if we were to assume the Netziv's conditions to *aveira lishma*, we run into difficulty when applying it to the case of saving the many at the expense of the one. A Responsum in the name of the Rama addresses such an application and, noting that the Marauders Case forbids saving the many at the expense of the one, forbids applying *aveira lishma* to save the many at the expense of the one.[58] Of course, we could use R. Medan's understanding that the Marauders Case doesn't apply in peacetime, but perhaps our proposal has become too maverick.

Between Robots and People

Before we jump to conclusions and apply the above understanding to the programming of the autonomous vehicle, we must ask if there is not some difference between a human driver and a digital driver. Initially, most would be inclined to say that there is no difference, for a computer simply automates what a human does. Upon deeper investigation, however, a number of differences that could be of significant ethical import can be discerned. Let us examine three levels of the autonomous vehicle which distinguish it from a human driver: the processor, the programmer, the system.

---

[56] Nefesh HaHaim 1:22. Similarly R. Kook, who also notes that today people unwittingly perform an *aveira lishma* and it is best that they do so unwittingly rather than knowingly sinning (Arpilei Tohar p. 15, quoted in R. A. Shavit, "Etgar HaDor", *Tzohar* 19 (Summer 5764) [Hebrew]).

[57] Responsa Noda Beyehuda (Tanina), YD 161.

[58] Responsa HaRama #11 quoted in N. Rakover, *Ends that Justify the Means* (Jerusalem: The Library of Jewish Law, 2000) [Hebrew], pp. 186-188. See ibid., p. 175 regarding the doubt of authorship. As an important side note, while Thomson initially wrote that sacrificing the few to save the many was a permissible wrong, she later retracted ("Turning the Trolley," *Philosophy & Public Affairs*, Fall 2008, Vol. 36, No. 4), thus coming to the same conclusion that we have arrived at here – i.e., the inapplicability of *aveira lishma*.

Processor[59]

At the heart of the decision-making process in the autonomous vehicle is what can be referred to as a central processing unit (CPU). In simple terms, the processor has inputs (e.g., number of people in my lane, number of people in adjacent lane), based on which it executes a program (e.g., move to adjacent lane if people in my lane), and drives outputs (e.g., turn steering wheel right).

A simplified program to handle the Trolley Problem according to utilitarian considerations (i.e., always save the many) would simply check which of the two lanes has fewer people and steer into that lane. The code could look something like this:[60]

```
Trolley_Routine (MyLane, AdjacentLane, Steering)
Begin
  If (MyLane.Count > AdjacentLane.Count) // which lane has more people?
    Then Steering <= AdjacentLane.Number; // save many at expense of few
    Else NULL; // Steering stays as MyLane.Number
End // Trolley_Routine
```

Here we see that the code causes the car to make an active move from the current lane (i.e., MyLane) into the adjacent lane. This would clearly appear to parallel the human driver making an active decision to turn the steering wheel and kill the person(s) in the adjacent lane (*kum v'aseh*).

Conversely, a program written according to the deontological approach (e.g., never sacrifice the few) would only change lanes if the adjacent lane is completely free. The code could look something like this:

```
Trolley_Routine (AdjacentLane, Steering)
Begin
```

[59] I thank Aviad Kipnis at Mobileye for this idea and pursuant discussions.

[60] The following discussion assumes hard-coded algorithms as opposed to a machine learning implementation. The outcomes of the discussion are pertinent regardless of implementation, though it should be pointed out that Mobileye intends to hard-code all safety related routines such that, while driving trajectories will be determined by machine learning, there will be a safety "governor" that is hard coded (See J. Yoshida, "Can Mobileye Validate 'True Redundancy'?", *EE Times* (5/22/2018), https://www.eetimes.com/document.asp?doc_id=1333308. See also Shai Shalev, https://www.youtube.com/watch?v=FovLsAFiIJU, starting at minute 26:00).

```
    If (AdjacentLane.Count == 0) // is adjacent lane empty?
      Then Steering <= AdjacentLane.Number; // save many at no expense
      Else NULL; // Steering stays as MyLane.Number
  End // Trolley_Routine
```

Here we see that if the adjacent lane is not free, the car simply stays in its own lane, the processor doing nothing. The fact that the code says "Else NULL" is just convention to allow you to see that no active change is made to the system. Accordingly, it would seem that the processor does nothing, and we can consider the car driving straight into the people in its current lane as passive killing (*shev v'al taaseh*).

However, if we dig a bit deeper, this high-level code is not what will be executed by the processor. Rather, this code will be translated into low-level assembly code which contains the actual processor instructions. Our deontological routine could look like this:

```
TROLLEY  LDA ADJLN   load register A with number of people in adjacent lane
         CMP A,0     check if adjacent lane is empty
         BNZ AIRBG   if adjacent lane not empty, branch to airbag routine
         LDB 1       load register B (that holds lane number) with 1 to change lanes
         JMP CTRL    jump to general controller routine
AIRBG    LDC 1       load register C (that holds airbag state) with 1 to inflate
         ...
```

What can be noted from the above assembly code, even by a layman, is that there is never a time when the processor is not executing an operation. At the processor level, there is no such thing as "passive" (*shev v'al taaseh*) since it is always actively executing instructions (and always updating its program counter). This being the case, it could be said that the processor will either actively decide to drive over the people in its current lane or actively decide to drive over the people in the adjacent lane.[61] Given this choice, clearly it should opt for killing as few people as possible.

This approach, however, is not without questions. When we look at the actions available to an agent in the Trolley Problem, we consider that he either actively performs an act (*kum v'aseh* – i.e., turns the steering wheel) or he remains passive (*shev v'al taaseh*). Can the execution of code in a processor really be considered parallel to such action? Is it not more akin to the decision-making process in the human mind? And if the human driver is only accountable for his act of turning the steering wheel and not the thoughts that led to it,

[61] R. Y. Medan, in a personal correspondence with the author, holds this to be a valid approach.

should we not, then, only consider the actuation of the steering wheel (or lack thereof) and not the execution of the code that led to it?

Programmer[62]

Going up a level from the processor, we arrive at the programmer himself. The programmer, while writing code to direct the car, is quintessentially different than the driver of a car. For, while the driver of the car is confronted with the Trolley Problem in real time, the programmer is confronted with it only in theory. In the case of a human driver, he will have been driving for some time after which he becomes faced with the life-and-death decision to either passively run over, say, five people in his current lane, or actively switch lanes and run over, say, one person. In contradistinction, the programmer writing his code does not arrive at a situation after driving in a specific lane such that he is faced with a choice to remain passive and continue in that lane or become active and change lanes. True, he is writing code for such a situation, but he himself is not confronted with a situation that has such a history. He is writing a simple if-then-else statement (as shown above).[63] The decision is being made at the time of writing the code, and consequently, it does not consist of passive and active alternatives. For the programmer making the decision when writing his code, he is faced with two equally *active* alternatives: write code to kill the many, or write code to kill the few. Clearly the decision should be to minimize deaths.

This approach, too, leads to questions. Can we really consider the decision made by the programmer in isolation of how it will come to be executed in real time? And, echoing the argument we made about the processor, if the human driver is only accountable for his act of turning the steering wheel and not the thoughts that led to it, should we not, then, only consider the actuation of the steering wheel (or lack thereof) and not the writing of the code that led to it?

System[64]

[62] I thank R. Haim Shahor at Mobileye for this idea and pursuant discussions.

[63] See above fn. 60.

[64] I thank R. Yisrael Malka at R. Weiss' Institute for this idea and pursuant discussions.

In contrast to the "processor" approach and the "programmer" approach, the system approach, while incorporating the fact that the programming is done before the dilemma is reached, seeks to address the dilemma in real time – i.e., at the time of the actuation of the steering wheel.

For the sake of clarity, let us restate the problem. There is an ideal that says, "strive to minimize losses." However, this ideal does not override the ideal that prohibits "killing with one's own hands" (*hariga beyadayim*). The Hazon Ish, as we learned, grappled with this problem using the example of the divert-able missile. He concluded that though one, in diverting the missile away from the multitude, would fulfill the ideal of minimizing losses, that same act of diverting the missile would inherently entail "killing with one's own hands" when the missile falls on the individual. This is true whether a human being is diverting a missile or driving a car. There is, however, a difference to consider when a computer is driving the car. R. Josh Flug offers the following approach:

> The Hazon Ish questioned the moral legitimacy of diverting the missile when the individual is actually standing in harm's way, for this we must say is like "killing with one's own hands." However, those who program an autonomous vehicle are only doing so to save the lives of the many and the programming does not take place at the time of the dilemma. Therefore, it seems more reasonable to say that this is not considered "killing with one's own hands" according to the Hazon Ish.[65]

R. Flug here raises two distinctions to eliminate "killing with one's own hands" and thus allow saving the many at the expense of the individual. One, the *modus operandi* of the system is "to save the lives of the many"; and two, the programming of the system – i.e., the act of choosing the individual instead of the many – is not made when "the individual is actually standing in harm's way."

This approach is refined by the rabbis who wrote the above-mentioned *Halachic, Ethical and Governmental Challenges in the Development of the Autonomous Vehicle*. They explain that when programming the system according to the *modus operandi* of "saving lives," we never enter into the question of killing or of judging between individuals, because we never look at

[65] R. J. Flug, "Pikuah Nefesh Matters Relating to Self-Driving Cars," *Yadrim* (Nisan 5777) [Hebrew]. R. Y. Medan, in a personal correspondence, makes this same assertion.

actual people. We are only looking at "statistical man" and trying to increase his chances of staying alive – that is, *everyone's* chances of staying alive.[66] When the system decides to hit the one and not the many, it is not deciding to kill, it is deciding to save. "Programming the car in a neutral environment [i.e., free of the exigencies of the road] and giving preference to the safety of the many cannot be thought of as a directive to murder but as an expression of the imperative to save lives."[67] Accordingly, the programming of the autonomous vehicle to save the many fulfills the Hazon Ish's criteria for framing a case as "saving": a) the individual is not sacrificed as a means but only incidentally, and b) his death is not considered *hariga beyadayim*.[68]

The Tunnel Problem

And what about the Tunnel Problem? Having noted a difference between the human driver and the digital driver with respect to the Trolley Problem, we must ask if there is not such a difference with respect to the Tunnel Problem. As you will recall, the Tunnel Problem pits the driver against the pedestrian in the road. Assuming that the pedestrian is in the road by all legal rights, and assuming the more stringent case that the car must actively turn into the curve, we learned that the human driver must passively sacrifice himself to avoid killing the pedestrian. Is there a difference between a human driver and a digital one? The authors of *Halachic, Ethical and Governmental Challenges in the Development of the Autonomous Vehicle* indeed note a difference:

> It is possible that an autonomous vehicle can be programmed to provide the greatest level of safety for its occupants, without taking into account that other people might be injured as a result. For the goal of such programming is not to injure anyone but simply to provide safety. The basic instruction to the system, "always stay in your lane and never drive off a cliff," is indisputably legitimate. The fact that this instruction ignores the possibility that a person may appear in the road does not make the car's failure to drive off a cliff to save the pedestrian an act of

[66] Malka, p. 181.

[67] Malka, p. 158.

[68] It should be noted that not everyone agrees with this logic. In "The Use and Abuse of the Trolley Problem" (*Ethics of Artificial Intelligence*, Oxford University Press, 2020), ethicist Frances Kamm writes explicitly that programming does not change the ethics – i.e., there is no difference between the digital driver and the human one.

> murder. When the system turns the steering wheel to follow the curve of the road, it does not do so as a specific act with relation to the current situation (including injuring a pedestrian) but rather as part of its overall design to protect lives and drive safely. No individual has the right to sue against a design that is fundamentally made to promote safety and not to injure.[69]

The argument being proposed is that a system designed for safety need not be concerned with injuring others as long as it is following the laws of safe driving. Furthermore, such a system that does kill an individual cannot be considered to have murdered him, since the action of a system programmed for safety, as explained before, is not considered *hariga beyadayim*. On the one hand, we can take comfort in the fact that the autonomous vehicle can legitimately save the passengers and not demand their self-sacrifice.[70] On the other hand, is it really sufficient to ignore outside surroundings just because murder is not involved?

Perhaps this "greatest level of safety for its occupants" approach can be supported by noting another difference between the human driven car and the autonomous vehicle: society. That is, when a human is driving his car, the dilemma takes place at the level of the individual, whereas when programming the autonomous vehicle, the dilemma is approached as one concerning society at large. At the societal level, halacha takes a much more cautious approach to danger, always seeking to ensure the general safety of the public.[71]

Accordingly, when programming the autonomous vehicle, we are not programming it for an individual *qua* individual, but for the individual as part of society as a whole. "From this global perspective … perhaps, when programming the autonomous vehicle, it should never be allowed to enter into situations that will endanger the occupants of the vehicle."[72] That is, looking at the autonomous vehicle from the societal level, it is incumbent upon the

---

[69] Malka, p. 178.

[70] I say, "take comfort," as studies show that, by and large, people would only buy an autonomous vehicle that protects the passengers (e.g., Jean-François Bonnefon, et al, "The social dilemma of autonomous vehicles", *Science*, Vol. 352, [June 24, 2016]).

[71] See Rakover, *Sacrifice*, Ch. 6. See also Malka, Sec. "Societal Approach to Danger," pp. 139-145.

[72] Malka, p. 145.

program to take every precaution to ensure the safety of the individuals entrusted to its care – i.e., the occupants of the car.

On the other hand, it could be argued that the perspective of "always seeking to ensure the general safety of the public" means that it is incumbent upon the program to take every precaution to ensure the safety of the individuals entrusted to its care – i.e., all members of society at large, regardless of whether they are inside the car or out. This approach would argue for utilitarianism, always seeking to minimize losses, whether that means saving the occupants of the car when they are more numerous than those on the road, or saving the pedestrians, for example, when there are more of them than occupants inside the car. However, in the case of equal numbers of passengers versus non-passengers, it would seem the dictum laid down by Rabbeinu Yona should hold sway – i.e., "one must remain passive when it is impossible to resolve whose blood is redder." The car would have to simply remain on course, consequences be what they may.

The Altruist

Finally, one solution rises above the ambiguities raised by the previous approaches: the altruistic solution. As we saw in the story of the Lulianus and Papus, who voluntarily sacrificed themselves to save the city, there is room to say that one can voluntarily give up one's own life to save others. The Tzitz Eliezer (15:70) limits the permit of self-sacrifice to cases in which the individual sacrifices himself for the "*klal*" – i.e., a large group, or "*klal Yisrael*" – i.e., a very large group.[73] Nevertheless, R. Kook is far more liberal in his permit of self-sacrifice.

> It can be said that even the Rambam[74] would agree that, in the case of spilling blood, if one wants to sacrifice himself for the sake of his friend he is permitted. Furthermore, it can be said that R. Akiva, who said "your life takes precedence,"

[73] See Tzitz Eliezer (17:72:12) according to Yeshuot Yaakov (YD 157:1).

[74] The Rambam writes: "Anyone for whom the rule is that [due to his situation] he is to violate a commandment rather than sacrifice his life, yet he sacrificed himself and didn't violate the sin – has forfeited his life" (Hil. Yesodei HaTorah 5:4). Accordingly, the Rambam does not accept that one can sacrifice himself whenever he deems appropriate. R. Kook, however, believes that here, in the extreme case when a life is on the line, the Rambam would agree to allow self-sacrifice. See Rakover, *Sacrifice*, p. 85-88, who gives this explanation but raises doubts if this is really the Rambam's opinion.

> came only to invalidate Ben Petora's position (that it would be best for both to drink and die). However, if one wanted to give [the water] to his friend because the life of his friend is dearer to him than his own life, in cases like this it can be said that there is no prohibition, even for the case of one saving one [let alone for one saving many].[75]

Consequently, if it were possible for a driver, human or digital, to avoid killing anyone but himself, such would be an acceptable, perhaps even praiseworthy, act.[76]

Summary

For a human driver confronted with the Trolley Problem, the weight of halachic authority rests squarely on the side of passively driving into the many to avoid actively killing even a single human being. There is room to say that, in light of the value of "striving to minimize losses," one could apply the rule of *aveira lishma* and swerve away from the many to kill the individual.

As for a human driver confronted with the Tunnel Problem, assuming the pedestrian under threat is within his legal right to be in the road, then, if the lane is straight, one may passively run over him. Conversely, if any active change is required of the driver, halacha requires the driver to passively sacrifice himself rather than actively run over the pedestrian. These conclusions also apply regardless of numbers, i.e., more people in the car or more people on the road. So if a car filled with passengers were heading toward a pedestrian around a bend, the driver would nevertheless be obligated to remain passive, causing the deaths of those in his vehicle to avoid actively driving over the pedestrian.[77] Conversely, if there were

---

[75] R. Kook, Mishpatei Cohen #143. Similarly, R. Shaul Yisraeli, permits self-sacrifice in that Ben Petora's position is not prohibited, but simply not commanded (*Amud HaYemani*, Machon HaTorah VeHaMedina, Jerusalem: Eretz Hemda, 1992 [1966], p.138: http://www.eretzhemdah.org/Data/UploadedFiles/FtpUserFiles/ravIsraeli/books/amudHayemini.pdf).

[76] On a technical note, to allow the autonomous vehicle to opt for altruism demands that the car owner be given the ability to toggle a switch that selects "altruist mode," for clearly no other agent can volunteer one to be an altruist.

[77] This is due to our inability to resolve "whose blood is redder" even in a one-against-many case (See fn. 47). And in this case, even an attempt to save the many by applying *aveira lishma* would be to no avail since the Netziv's first condition to allow such an act is that it be of no personal enjoyment/benefit.

many people standing in the middle of a straight road and one person in the car, the driver would not be obligated to commit suicide to save the many, but could passively drive the car straight into the people.[78]

For the autonomous vehicle, however, different conclusions can be drawn. Regarding the Trolley Problem, an argument can be made for always saving the many, since there is no *hariga beyadayim*. Regarding the Tunnel Problem, an argument can be made to always save the passengers by programming the car to simply abide by the rules of the road. But perhaps these interpersonal-level arguments are trumped by approaching the dilemmas from the societal level which demands extra precaution in safeguarding public welfare and would consequently call for the autonomous vehicle to be programmed to always provide the utmost safety for the public at large – i.e., program according to utilitarianism.

In any of the above dilemmas, whether the driver is human or digital, taking the altruistic decision of self-sacrifice, if available, is always open. What is not open is deciding not to decide. That is, due to the difficult nature of these ethical dilemmas, some suggest that the car should be programmed to decide using a random number generator – what amounts, in layman's terms, to flipping a coin. Someone mentioned to me that this puts the decision in God's hands. Without addressing the veracity of that statement, I would argue that flipping a coin is an abandonment of the very responsibility given to us by God, since moral decisions are "not in Heaven" (Baba Metzia 59b).[79] Furthermore, the Hazon Ish invalidates such an approach based on the Marauders Case, writing, "If running a lottery [to choose an individual to give to the marauders] were an option, why did the Yerushalmi teach, 'let

[78] Rav Kook (Mishpatei Cohen #143) discusses this case, wondering if the Tosafot (who allow one to be passively thrown on a single baby) would allow one to be passively thrown onto several babies. He conjectures that here the individual might be required to passively sacrifice his own life to save the many. Nevertheless, he concludes: "this needs investigation." Interesting though this is, it does not apply to driving scenarios; since if one were passively driving into several "babies," one would have no way to save them by giving his life passively, but only by actively committing suicide – an act which, according to majority opinion, is not obligatory.

[79] Also, "All is in the hands of heaven, except for the fear of heaven" (Ber. 33b, esp. Rashi ad loc). Ethicist Patrick Lin puts it like this: "Making decisions randomly, then, evades that responsibility. Instead of thoughtful decisions, they are thoughtless, and this may be worse than reflexive human judgments that lead to bad outcomes" ("The Robot Car Of Tomorrow May Just Be Programmed To Hit You," *Wired*, https://www.wired.com/2014/05/the-robot-car-of-tomorrow-might-just-be-programmed-to-hit-you/). Similarly Walter Wurzberger, *Ethics of Responsibility* (Philadelphia, PA: JPS, 1994), p.91.

everyone die and hand over no one,' when it could have taught, 'run a lottery and hand over the individual upon whom the lot fell'?!"[80]

As mentioned at the outset, the purpose of this article is educational, seeking to bring to the fore various dilemmas along with the sources and arguments used to think about their solutions. It is certainly not meant to tell programmers how to write code. Indeed, a great discussion revolves around who should decide the ethics of how an autonomous vehicle should be coded: the programmer, the technology provider, the car manufacturer, the car owner, the government regulator, or someone else? It seems to me that the decision lies squarely in the hands of government regulators who represent the interest of the people and who are guided by professional ethicists. All the other parties mentioned have an agenda that does not necessarily align with ethical interests: e.g., the programmer wants to keep his code simple, the technology provider wants to minimize costs, the car manufacturer wants to maximize sales, the car owner wants to save himself.[81] Accordingly, it is my hope that this paper will provide food for thought to ethicists making these fateful decisions. Finally, given the grave consequences of this subject matter which have caused me to literally lose sleep for weeks, I conclude with the words of those far greater than I who have weighed in on this issue: "This needs a lot of investigation."

Postscript

I would be remiss not put this whole discussion into proper context. That is, though these ethical dilemmas are very real, the actual occurrence of them will be very unlikely. So while the discussion is indispensable since autonomous vehicles will have to be programmed to account for them, in reality the algorithms will rarely be executed. This point cannot be emphasized enough because the public needs to realize that autonomous vehicles are being developed to save lives, not endanger them. Indeed, the sooner autonomous vehicles are on the road and human drivers removed therefrom, the sooner we can start saving literally millions of lives.

---

[80] Hazon Ish (San. #69). See also Yabia Omer Vol 6, HM #4.

[81] This conclusion seems to be understood, as can be seen in statements by the CEO of Mercedes ("Mercedes-Benz's Self-Driving Cars Would Choose Passenger Lives Over Bystanders," *Fortune*, October 15, 2016) and Mobileye (personal conversations).