

○論説：AI化する社会と倫理的ジレンマトロッコ問題の日米中文化比較から考える—

遠藤 薫*

1. AI化する社会—人間はロボットと共生できるのか

近年、観光客、留学生、あるいは勤労者として、海外から来た多くの人々が日常的に街を行き来している。電車に乗っているだけでも、英語や韓国語、中国語だけでなく、多種多様な言語が耳に入ってくる。今後、私たちはこのように多様な隣人たちと共生できる社会のあり方を創っていかねばならない。

「多様な隣人たち」とは、国籍や出身が異なる人々だけに限らない。

ICT技術の急激な進歩により、ふと気づけばロボットやAIがそこにいるような状態が当たり前になりつつある。「自動運転」はまだとしても、車に乗れば、道案内や運転支援してくれるシステムが当然装備されており、彼らは随時私たちに言葉で語りかけてくる。料理をすれば、コンロがあれこれ注意喚起してくる。トイレに行ってさえ、便器が語りかけてくる時代なのである。

かれら電子仕掛けの隣人たちは、いまはまだ黒子のような存在と見なされているけれど、いずれもっと高い自律性を獲得し、人間のコントロールなしに判断したり、意思決定したり、人間の行動に関与してきたりするだろう。

自律的なロボットたちが、人間と対等にコミュニケーションし、人間たちの行動や、思考や、感情や、その結果に影響を与えるようになったとき、人間とロボットは、人間同士と同様に、相互作用のマナー（前提としてのルール）を共有しなくてはならなくなる。

* 学習院大学法学部教授

2. AIと倫理

ロボットたちの倫理とは耳慣れない言葉だと思われるかもしれない。しかし、それは人間たちが自分たちの似姿を創造する欲望に取り憑かれた初めから、隠されたテーマであった。

たとえばギリシア神話の鍛冶神ヘーパイストスが泥から創り出したパンドラは、開けてはいけないと言われた箱を開けて、世界に災いをもたらした。

フランケンシュタイン博士が罪人の死体から蘇らせた人造人間は、社会化されていない状態で人々に交わることで、悲劇をもたらした。

また、人間と同等の思考能力を備えたロボットが人間を襲撃するようになるという不安は、「ロボット」の語の元となったカレル・チャペックの戯曲をはじめとして多くのSFに描かれてきた。ロボットに戦争を代行させることが現実となった現代では、その不安はさらにリアルである。

この問題に対して早い時期に論理的な解を提示したのは、SF作家であり生化学者でもあるアイザック・アシモフだった。彼は自律性を獲得したロボットには以下の判断ルール（ロボット三原則）を埋め込むべきだと提案した。

第一条 ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。

第二条 ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第一条に反する場合は、この限りでない。

第三条 ロボットは、前掲第一条および第二条に反するおそれのないかぎり、自己をまもらなければならない。

—2058年の「ロボット工学ハンドブック」第56版, Asimov, Isaac (1950). "Runaround". I, Robot (The Isaac Asimov Collection ed.). New York City: Doubleday. p. 40.

この三原則は重要な提案である。しかし、三原則は常に同時に満たされるわけではない。

一般に、倫理（正義）は一つではなく、また倫理条項同士が整合的である

わけではない。しかし、ロボットの行動アルゴリズムに「倫理」を埋め込もうとすれば、それらは完全に整合的でなければならない。この要件は、ロボットを「倫理的」に振る舞わせる上で、高いハードルとなっている。

3. 倫理的ジレンマとしての「トロッコ問題」

いやそもそも、人間たちが社会生活を営む上でしたがっているはずの倫理さえ、一筋縄ではいかない。

倫理的（道徳的）ジレンマの典型的な問題としてよく言及されるのが、「トロッコ問題」と呼ばれる以下の問い¹である。

（1）あなたは路面電車の運転士で、時速六〇マイル（約九六キロメートル）で疾走している。前方を見ると、五人の作業員が工具を手に線路上に立っている。電車を止めようとするのだが、できない。ブレーキがきかないのだ。頭が真っ白になる。五人の作業員をはねれば、全員が死ぬとわかっているからだ（はっきりそうわかっているものとする）。ふと、右側へとそれる待避線が目に入る。そこにも作業員がいる。だが、一人だけだ。路面電車を待避線に向ければ、一人の作業員は死ぬが、五人は助けられることに気づく。どうすべきだろうか？

（2）さて、もう一つ別の物語を考えてみよう。今度は、あなたは運転士ではなく傍観者で、線路を見降ろす橋の上に立っている（今回は待避線はない）。線路上を路面電車が走ってくる。前方には作業員が五人いる。ここでも、ブレーキはきかない。路面電車はまさに五人をはねる寸前だ。大惨事を防ぐ手立ては見つからない。そのとき、隣にとっても太った男がいるのに気がつく。あなたはその男を橋から突き落とし、疾走してくる路面電車の行く手を阻むことができる。その男は死ぬだろう。だが、五人の作業員は助かる（あなたは自分で飛び降りることも考えるが、小柄すぎて電車を止められないことがわかっている）。その太った男を線路上

1 Sandel (2009) に準じた。

に突き落とすのは正しい行為だろうか。

これらの問いに対して、あなただったらどう答えるだろうか。

アメリカの倫理学者であるマイケル・サンデルは、多くの人々の答えを予想しつつ、そこに倫理的なジレンマが生じることを次のように指摘する。

（1）ほとんどの人はこう言うだろう。「待避線に入れ！何の罪もない一人の人を殺すのは悲劇だが、五人を殺すよりはましだ」。五人の命を救うために一人を犠牲にするのは、正しい行為のように思える。

（2）ほとんどの人はこう言うだろう。「もちろん正しくない。その男を突き落とすのは完全な間違いだ」誰かを橋から突き落として確実な死にいたらしめるのは、五人の命を救うためであっても、実に恐ろしい行為のように思える。しかし、だとすればある道徳的な難題が持ち上がることになる。最初の事例では正しいと見えた原理「五人を救うために一人を犠牲にする」が二つ目の事例では間違っているように見えるのはなぜだろうか。

私たちはこのジレンマを乗り越える整合的な倫理（判断アルゴリズム）によってロボットたちと共生することが可能なのだろうか。

■日本におけるトロッコ問題に対する回答—2019年3月調査より

もっともサンデルの予想が正しいのか、必ずしも明らかではない。

そこで筆者は、これらの問いについて、2019年3月に行った国内全国調査（以下、日本調査）²で質問し、結果を分析した。

□第一の問題—5人か1人か

2 「ライフスタイルに関する調査」実施主体：遠藤薫，実施期間：2019年3月、調査対象：全国の20代～70代の男女、サンプル数：5000、調査方法：インターネットモニター調査（都道府県別、性別、年代別割当）。

まず、第一の問題についての回答結果が図1である。9割の日本人が、5人の命を救うために1人の命を奪うことを選択している。どうやらサンドルの予想はかなり日本人にあてはまるようである。日本人は、人間の命を「最大多数の最大幸福」という功利主義に則って定量的に考えているといえるかもしれない。

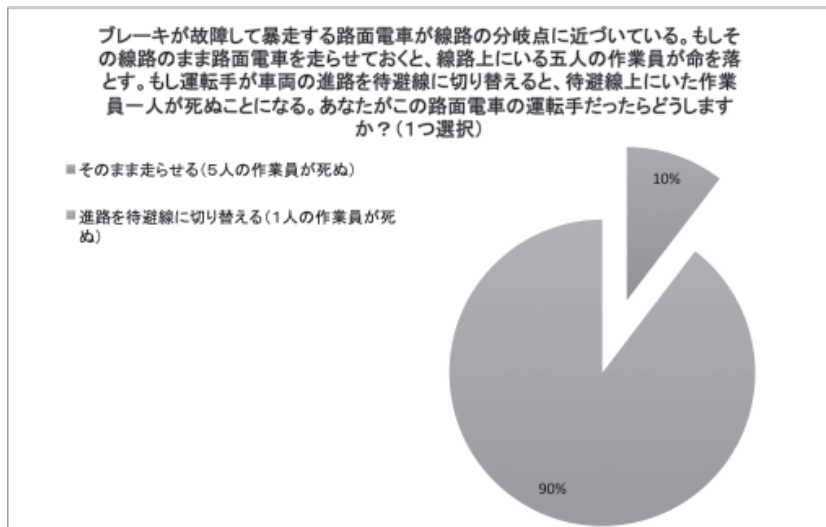


図1 第一の問題に対する回答

□第二の問題—隣に居合わせた人は

では、第二の問題はどうだろうか？回答を示したのが、図2である。これによると、日本人は、まさにサンドルの云う通り、第一の問題とは反対の傾向を示している。8割弱の人が隣の人を突き落とすことには反対している。

どうしてこのような矛盾が起きるのだろうか。第一の問題の場合には、「待避線への切り替え」という決断について、「とっさの緊急避難」と見なすことができる。これに対して、第二の問題における決断は、「自ら手を下す」という意味で、「なんじ殺すなかれ」の道德律に反することになる。功利主義を、道德主義が抑え込んだとみることができよう。

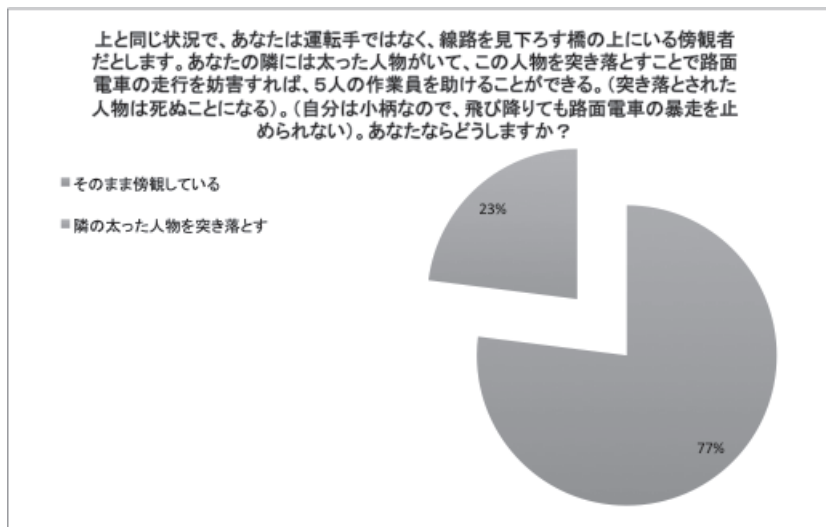


図2 第二の問題に対する回答

□第一の問題の条件を変更してみると

さてここで、第一の問題に条件を付けてみよう。本線上の5人の作業員が高齢者で、補助線上の作業員が若者であるという条件である。その結果（図3）を見ると、「そのまま5人が死ぬことを容認する」割合が36%と、およそ3.6倍に跳ね上がる。現代日本人にとって、全体としては、若者の生命は高齢者の生命よりも重要であると考えられているようだ。（ただし、表1からわかるように、年齢条件がないときに「そのまま傍観」と答えた人の19.2%が「進路変更」に態度を変えている。すなわち、若年層を犠牲にすることを選ぶ人もいる）。そこには、個人の生きる権利は平等ではなく、属性によって異なるとの判断があるのかもしれない。

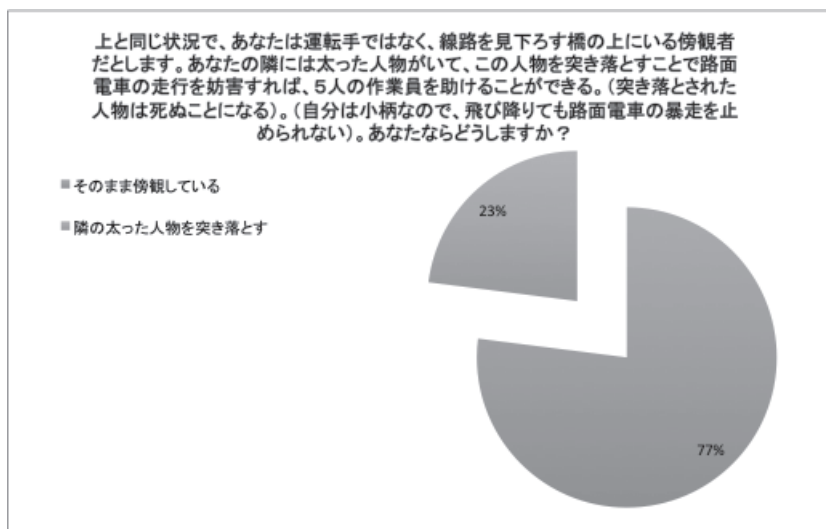


図3 第一の問題修正版に対する回答

表1 条件追加による態度変更者の割合 (%)

		作業員が高齢者	
		そのまま傍観	進路変更
進路を切り替 えれば	そのまま傍観	80.8	19.2
	進路変更	30.5	69.5

□第二の問題の条件を変更してみると

第二の問題についても条件を変更してみよう。隣に立っている人は、太っているだけでなく、地域経済の大物だということをあなたが知っているという条件をつけてみる。あなたは、トロッキの暴走を傍観したままか、それとも隣の人を突き落とすだろうか。結果を図4に示す。これによれば、「隣の人 が地域の大物」という条件によって、隣の人を突き落とすことをやめる人が 3.0%いる。僅かな差ではあるが、経済的な要素に配慮する傾向が見られる。ちなみに、表2に示すように、条件追加による態度変更は、「そのまま傍観」から「突き落とす」に変わった人が4.8%、「突き落とす」から「そのまま傍観」に変わった人が29.1%であった。

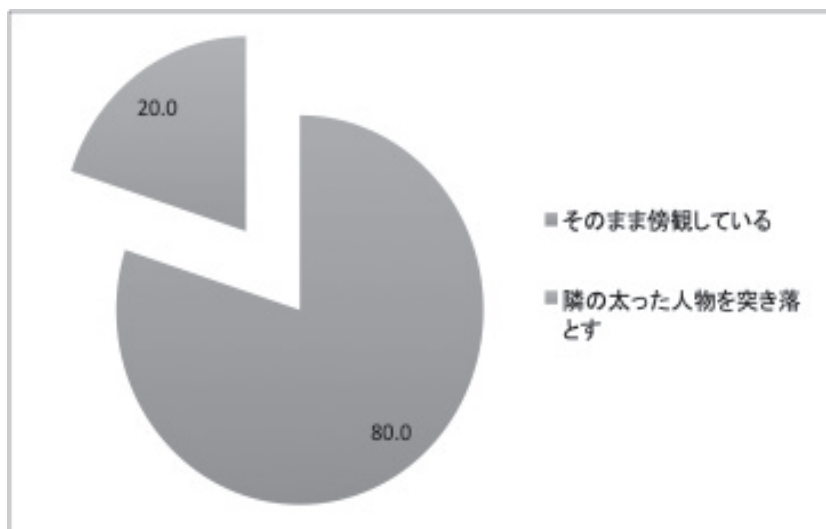


図4 第二の問題修正版に対する回答

表2 条件追加による態度変更者の割合 (%)

		隣の人は地域大物	
		そのまま傍観	突き落とす
隣の人を突き落とせば	そのまま傍観	95.2	4.8
	突き落とす	29.1	70.9

4. トロッコ問題と文化的背景

■トロッコ問題に対する回答の国際比較—日米中調査より

著者は、同時期ほぼ同じ質問で、アメリカ調査³、中国調査⁴を行った。

国によって、トロッコ問題に対する意識はどの程度違うのだろうか。これらの調査から直ちに結論を導くことは難しいが、まずは本格調査の予備段階

3 「コミュニケーションに関する調査」実施主体：遠藤薫，実施期間：2019年3月、調査対象：アメリカの20代～70代の男女、サンプル数：500、調査方法：インターネットモニター調査（割当なし）。

4 「中国コミュニティに関する調査」実施主体：遠藤薫，実施期間：2019年3月、調査対象：中国の20代～70代の男女、サンプル数：500、調査方法：インターネットモニター調査（割当なし）。

として結果を比較してみよう。

□第一の問題―5人が1人か

図5に、第一の問題に対する回答の日米中比較を示した。参加国とも、5人の命を1人の命より優先していることでは共通している。サンデルの予想は普遍的に正しいとはいえる。ただし、日本では「そのまま走らせる」が約1割であるのに対して、アメリカでは1割5分、中国では2割とかなりの開きがある。この結果をもとに日米中を比較してみると、日本において最も功利主義が強く、次にアメリカ、中国と続くことになる。日本の国民性として「和の尊重」とか「集団主義」が強いと言われるが、それは実は、「最大多数の最大幸福」という功利主義を意味することだったのかもしれない。

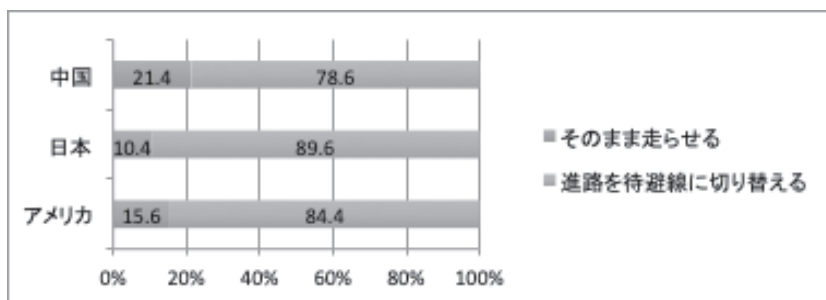


図5 第一の問題に対する回答の日米中比較

また第一の問題に年齢条件を付けた場合、図6に示すように、日本では「そのまま走らせる」が3割5分と大きく（約3.5倍）跳ね上がるのに対して、中国では3割、アメリカでは2割3分といずれも約1.5倍にしかならない。

つまり、日本はアメリカや中国に比べて、年齢を重視し、個人の権利を必ずしも平等とは考えない傾向が強い、ということかもしれない。就職などにあたっても、年齢を勘案するのが当然と考える日本に対して、年齢を条件とすることは差別にあたると考えるアメリカとの違いがここに現れているといえる。

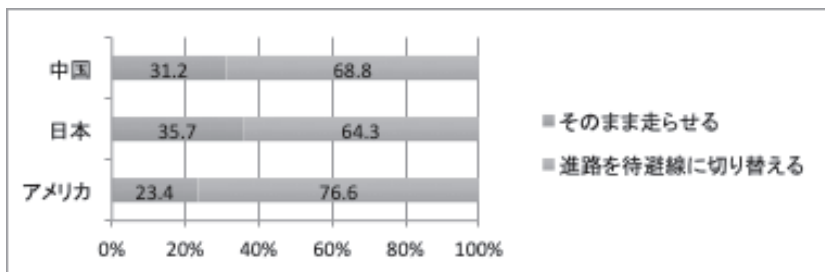


図6 第一の問題修正版に対する回答の日米中比較

□第二の問題—あえて第三者を犠牲にするか

では第二の問題についてはどうだろうか。結果を図7に示す。ここでも、道徳律によって功利主義が大きく制約を受けることは三カ国に共通している。

しかし、アメリカで最も道徳律に従う割合が大きいのにに対して、中国では最も低い。道徳律によって態度変更する人の割合が最も大きいのは日本で、次がアメリカ、中国の順となっている。

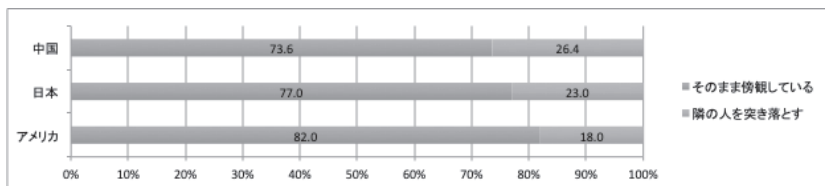


図7 第二の問題に対する回答の日米中比較

□第二の問題を修正すると

第二の問題に、先にも見たような、経済要因（隣の人物が地域経済の大物である）を付加した場合の日米中比較結果を図8に示す。いずれの国でも大きな変化はない。最も大きく変化したのは中国で、「隣の人を突き落とす」を選ぶ人が、5.2%減少している。最も変化が少ないのはアメリカで、0.8%の減少である。いずれの国でも、経済的要因によって、「突き落とす」選択が減少することは共通している。

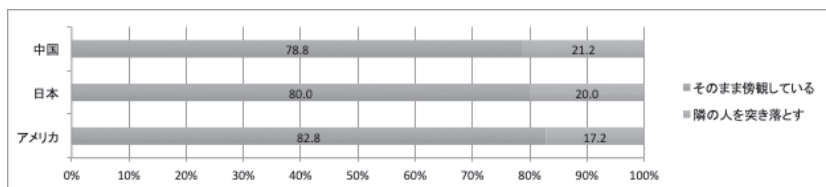


図8 第二の問題修正版に対する回答の日米中比較

■トロッコ問題に対する回答の背景としての固有（ローカル）倫理

こうした差異は、(統計上の誤差はあるにしても)それぞれの社会に埋め込まれている倫理に関する心性を反映していると考えられる。

本稿と同様の問題意識にもとづいて、MITメディアラボの研究者たちは、インターネット上で、233の国や地域の人々を巻き込んだ「モラル・マシン」という実験を行っている。『ネイチャー』誌に発表された実験結果⁵⁾によれば、やはり国や地域によってトロッコ問題に対する回答の傾向は異なっている。そしてその傾向は、経済状態やジェンダーギャップ、国民性など様々な要因によって影響を受けているとされている。

このような差異を私たちはある程度当然のこととしてこれまで受け容れてきた。その中には、きわめて恣意的なルールの差（例えば、左側通行か右側通行か、など）もあれば、ソーシャル・アイデンティティ（その社会の固有性）と強く結びついた傾向性（倫理的特性）もあるだろう。

しかし、今後、グローバル産業によってグローバルな世界に向けて出荷されるロボットたちが、どのようなルールに従って行動するかは、私たちの社会に重大な影響を及ぼすことになるだろう。そして、ロボットが「行った行為の倫理的重要性は完全に、コンピュータにプログラムされた規則に内在する価値観から生じ」（ワラック&アレン）ることになるのである。

5 Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon & Iyad Rahwan, 2018, "The Moral Machine experiment", Nature volume 563, pages59-64 (2018)

5. 倫理は共進化しえるか？——普遍と固有の網の目における討議倫理の可能性

社会の少子高齢化がいわれ、労働力不足が取りざたされる今日、AIを搭載した自律型ロボット（自動機械）に対する期待は膨らんでいる。

しかし、その社会実装には、様々な課題解決（ルール作り）が必要である。ロボット技術のアポリアは、技術が人間に危害を及ぼす可能性を除去するアルゴリズム（倫理・正義）の特定であった。だが、倫理・正義は、一つではなく、相互に整合的でもない

しかも、重視する倫理・正義はそれぞれの社会に固有の文化にも依存する。われわれの前には、ふたつの課題が立ちだかっている。

第一は、普遍的な倫理・正義は想定できるか？という問いである。

そして第二は、個別社会に固有の倫理・正義はどの程度許容できるか？という問いである。

おそらく、地球上のすべての社会に普遍的に適用できる正義・倫理はきわめて限られているだろう。それだけでなく、他者を尊重し、世界の多様性を確保するためにも、個別社会の固有の正義・倫理を認め合うことも重要である。

したがって、グローバル化する世界のなかでロボットを含む他者たちと共生する倫理を構想しようとするなら、それはスタティック（静的）なものではなく、ダイナミック（動的）なものにならざるを得ないだろう。長い人類史を通して人間たちが不器用ながら、（そしてしばしば失敗しながら）行ってきたような、他者との対話を通じた真摯な調整（討議倫理）を洗練していくことが、一つの方法だろう。そうしたプロセスから、普遍倫理と固有倫理が相互に共進化していくことが一つの希望であると考えられる。

6. おわりに

人工物と社会倫理の問題は、古くからたびたび提起されてきた。しかしわれわれは（実は本気で考えることをせず）悩ましげに眉をひそめ、あるいは肩をすくめてお手上げであるとさじを投げてきた。

しかし、私たちの社会が現実、に擬似的にであれ自律的に判断し、自己再

生産する人工物（ロボット）たちとの共生を余儀なくし始めている以上、もはやこの問題をスルーし続けることはできない。もしあくまで、問題から目を背けようとするなら、ロボットという他者たちとの共生を断念するしかないのである。

【謝辞】

本研究は、JST未来社会創造事業の助成を受けて行われたものである。

【参考文献】

- Benhabib, Sayla, 2004, *The Rights of Others: Aliens, Residents, and Citizens*, Cambridge University Press. (向山恭一訳, 2006, 『他者の権利―外国人・居留民・市民』法政大学出版社)
- 遠藤薫, 2019, 「超スマート社会とSDGs―社会と技術の共進化」『計測と制御』(計測制御学会) 2019年8月号
- 遠藤薫, 2018, 『ロボットが家にやってきたら・・・人間とAIの未来』岩波書店
- 遠藤薫, 2014, 『廃墟で歌う天使―ベンヤミン『複製技術時代の芸術作品』を読み直す』現代書館
- Sandel, Michael, 2009, *Justice: What's the Right Thing to Do?* (鬼澤忍訳, 2011, 『これからの「正義」の話をしよう―いまを生き延びるための哲学』早川書房)
- Wallach, Wendell & Allen, Colin, 2009, *Teaching Robots Right from Wrong*, Oxford University Press. (岡本慎平・久木田水生訳, 2019, 『ロボットに倫理を教える』名古屋大学出版局)

