Meta-surrogate decision making and artificial intelligence

Brian D Earp

How shall we decide for others who cannot (currently) decide for themselves? And who—or *what*, in the case of artificial intelligence—should make the decision? The present issue of the journal tackles several interrelated topics, many of them having to do with surrogate decision making. For example, the feature article by Jardas *et al*[1] explores the potential use of artificial intelligence (AI) to predict incapacitated patients' likely treatment preferences based on their sociodemographic characteristics, raising questions about the means by which we come to decide for others. And a clinical ethics round table led by Wilkinson and Pillay[2] examines the case of a premature baby on life support whose primary surrogate is herself incapacitated. Together, these examples force us to think more deeply about the meaning and significance of taken-for-granted concepts: respect for autonomy, substituted judgement, best interests. We'll consider the baby first and then turn to AI.

"Baby T" is a critically ill newborn delivered prematurely by emergency caesarean section. The mother had entered into a surrogacy arrangement with a same-sex male couple, the intended parents, who were to take over the baby's care after birth—just as soon as a formal parental order could be obtained through the court. Until then, the birth mother, who had used her own eggs to conceive Baby T along with sperm from an unidentified donor, would have legal and ethical responsibility to decide about the baby's care (also the right to keep Baby T if she chose). Unfortunately, she too was in critical condition, having fallen unconscious prior to delivery due to a sudden brain haemorrhage. She remained unconscious, and thus incapacitated, during a crucial period in which time-sensitive decisions about Baby T's care needed to be made, including whether to continue life support.

Given the mother's incapacity, who should determine Baby T's care? According to the analyses of Pillay *et al*[3] and Jackson *et al*,[4] the intended parents, although clearly both ethically and emotionally invested in these decisions, would not at that point have the *legal* authority to

make them. Instead, the spouse or civil partner of the birth mother would be the legal second parent (unless they had not consented to the surrogacy arrangement) until parenthood could be officially transferred to the intended parents through a court order or adoption. No second parent is mentioned in this case, and there isn't time to transfer parenting rights to the intended parents before key decisions need to be made. Although the commentators agreed that the couple should not be marginalised, but rather substantially included in discussions about Baby T's care,[5] the legal position seems to be that, in such a scenario, it is Baby T's doctors who would have the final say.

There is an interesting question here about the standard that should guide the doctors' decision making. In his commentary, Dominic Wilkinson[6] asks us to suppose that Baby T's prognosis is neither so poor that that treatment must cease, nor so good that it must continue. The decision, then, might be said to fall within what is sometimes called the "zone of parental discretion"[7] (for a critique, see[8]). According to this view, if the mother had had the capacity to decide, the medical team would have been obligated to follow her instructions (assuming she was adequately informed, and so on) regardless of whether they themselves agreed that the decision was in Baby T's best interests. Given that the mother did *not* have the capacity to decide, however, what should the doctors do?

Let us add a few more stipulations. Suppose the mother is unlikely to regain capacity any time soon, and the treatment required to keep Baby T alive is painful and invasive. Treating Baby T indefinitely while waiting for the mother to recover therefore isn't the obvious answer. Even the intended parents are split on what to do. The doctors need to decide whether to continue a painful treatment despite an unclear prognosis or withdraw treatment out of compassion for the baby's suffering. Should they

(1) try to infer what the mother would have decided—based on her values, wishes, cultural commitments, or religious

beliefs, for instance—and make a *substituted judgment* on her behalf, or

(2) simply do, directly, whatever they believe is in Baby T's *best interests*, whether or not they think it is what the mother herself would have decided?

The answer depends, in part, on how we conceive of the ethical basis for parental "proxy" decision making. There are two main schools of thought, one that is arguably more child-centred and one that is arguably more parent-centred, but we can start with common ground. First, it widely acknowledged that most parents love their children, deeply, and truly want what is best for them (that is, they have a maximally strong **motive of beneficence** toward their children). Moreover, parents usually are better positioned to *know* what is best for their children than just about anyone else (that is, they have **special epistemic access** to what is, in fact, in their child's best interest). So, for any decision that needs to be made about a child's treatment in a medical context, if the child is insufficiently autonomous to make their own decision, the parents should—barring exceptional circumstances—decide on their behalf.

There are two different ways of glossing this conclusion, however. The child-centred way suggests that, ultimately, the right thing to do is simply *whatever is in the child's best interests* (the best interests standard),[9] whereas, deferring to parental judgement just happens to be the most reliable general decision procedure for figuring out what that is (given **motive of beneficence** and **special epistemic access**). So, the parents should be deferred to.

The parent-centred way adds a premise: parents, on this view, have a fundamental right to make decisions about their children's upbringing, including their healthcare, in the context of wider family life and communal considerations; it is therefore wrong to interfere with, or override, their parenting decisions—even if those decisions are not necessarily in the child's best interests—unless the child is put at a significant risk of serious harm (the so-called harm principle).[10] However, this view continues, given **motive of beneficence**, most parents do not want to harm

their children, so there is no compelling reason to challenge this basic picture on grounds of children's welfare or rights.

The first, "best interests of the child" gloss is basically consequentialist, although tethered to the welfare interests of a focal individual: the child-patient. It says: whoever has the authority to decide about a child's treatment should weigh up the child-relative goods and bads of each feasible option, and choose the option that is all things-considered best for the child (or at least among the "good enough" options), given the child's particular welfare interests.

In the case of Baby T, the child's parent—the one who would usually have the authority to decide—is incapacitated. However, plausibly, she would not know any more about the child's *specific* welfare interests (vis-à-vis treatment options) than would Baby T's doctors, given that Baby T is a newborn who hasn't yet developed unique personal needs. Since the "defer to the parent" decision procedure is not available in this case, and the parent plausibly would not have **special epistemic access** anyway, the doctors should, according to this analysis, simply make their own informed judgement about what is best for Baby T.

The second, "parental rights" approach, by contrast, is more about respecting autonomy—parental autonomy. According to this perspective, parents' decisions are to be respected as such, irrespective of the likely consequences for child, unless the child is put at a significant risk of serious harm. In the case of Baby T, it has been stipulated that the decision to continue, or not to continue, life support are both within the zone of parental discretion. So, the correct thing to do, on this analysis, is to try to infer what Baby T's mother would have chosen—for example, based on her cultural values or religious beliefs—and make a substituted judgement on her behalf.[i]

Suppose that the hospital where Baby T is being treated has adopted a policy in line with the second approach: when a baby's mother is incapacitated and there is no second parent to decide—leaving time-sensitive, life or death decisions to the clinicians—they should not simply do what they think is in the best interests of the baby; rather, they should try to infer what the mother would decide (irrespective of the child's interests, but within the zone of parental discretion) and act accordingly.

However, suppose the clinicians don't know much about what Baby T's mother, in particular, would decide—they only have some general information about her demographic background. They know her age, gender, racial or ethnic categorization, city of residence, and perhaps the type of church she attends. There isn't enough time to try to bring in friends or family for special interviews. They need to make a substituted judgement as quickly as they can.

Perhaps they can fire up the Patient Preference Predictor (PPP)? In their feature article,[1] E.J. Jardas, David Wasserman, and David Wendler describe a proposed computer-based algorithm that would use machine learning (a type of artificial intelligence) to predict an incapacitated patient's treatment preferences based solely on their sociodemographic characteristics. Applied to the Baby T case, the preferences to be predicted would be slightly different: not those of an incapacitated patient regarding her own treatment, but rather, her preferences regarding the treatment of her non-competent child (a kind of meta-surrogate decision making). But let's simplify, going forward, and think about predicting only self-directed treatment preferences.

By drawing on existing correlations between past patients' treatment preferences and their sociodemographic characteristics, the PPP could, hypothetically, make predictions about current patients' preferences that were more accurate than the guesses of their real-life human surrogates. In fact, existing data suggest that a preliminary PPP prototype is already about as accurate as human surrogates,[12] so this is not an unreasonable hypothesis. Suppose it comes to pass. Now, a patient is incapacitated, there is no advance directive, there isn't time to reach out to family and friends; the doctors must decide about treatment.

Ordinarily, if they knew nothing in particular about a patient's preferences under such conditions, doctors would resort to a "best interests" standard and act accordingly. At first, this might seem quite different from the substituted judgement standard that is supposed to apply to once-competent patients who are currently incapacitated. According to that standard, the way to show respect someone who was previously autonomous, but who is now unable to make a treatment decision on their own behalf, is not to ask, "What do I or anyone else think is *best* for them?" but rather, "What would *they* decide for themselves in this situation?"

However, if "they" are essentially a black box, the best interests standard and the substituted judgement standard arguably amount to the same thing. It's like asking, "What would someone with no idiosyncratic preferences or desires—a fully informed, abstract, rational, self-interested person with no individuating features—choose for themselves if they were in this situation?" The answer is: "Whatever is in their best interests."

But the prospect of a PPP changes things. It invites us to fill in the "black box" and return to a more fully-fledged substituted judgement standard. By plugging in whatever limited information we have about the patient—their age, race, gender, and so on—we can make an empirical prediction about what the patient would, if autonomous, have in fact decided for themselves, over and above a rational "best interests" abstraction. And the prediction would be based on previously established correlations between those very same demographic variables and actual past patient preferences regarding treatment under similar conditions.

We are supposing that there isn't time to consult the patient's family or friends to find out more particular information. The doctors can either resort to a bland "best interests" test, or they can plug the patient's demographic information into PPP, which we are stipulating is known to be better, on average, at accurately predicting patient preferences than human surrogates. Should the doctors use the PPP?

Jardas *et al* consider a number of objections, according to which the PPP should not be used. One of them holds that, although there may be population-level statistical correlations between certain demographic features and associated treatment preferences, this is misleading at the individual level (that is, the level at which PPP-inspired treatment decision would actually be made). After all, one's group-level demographic features are not themselves the cause of one's individual-level preferences.[13]

True enough, say Jardas *et al*. However, the PPP does not assume that group-level demographic factors *cause* individual-level preferences. It simply harnesses those

---

[i]This appears to be Wilkinson's[6] view: "as with other situations where an adult lacks capacity, it may be possible to know enough about the surrogate mother's views and values to apply to the situation. It would be important to ask those who knew her well what he wishes were. Her partner or those close to her could indicate whether she has any relevant religious or other values, or whether she has ever expressed views about continuing intensive care in a setting where a child might have severe long-term disability." See also.[11]

group-level factors to make an empirical prediction about one's *likely* treatment preferences, above chance. Given that the alternative would be to make a nondescript "best interests" decision—one that is no more likely to be what *you*, in particular, would make than what any other random (rational, fully-informed, self-interested, etc.) person would make—isn't the PPP more respectful of your autonomy?

Another objection resists this move.[14] It holds that respecting someone's autonomy "is not simply a matter of treating them the ways they prefer to be treated. It is also important to make decisions for the right reasons, reasons the patient would endorse".[1] In response, Jardas *et al* suggest that there may be a trade-off, in certain cases, between respecting someone's autonomy in the sense of how they actually want their life to go (based, in turn, on on how they are treated) and honouring their assumed wishes for having surrogate decisions made for them according to a specific decision-making process (eg, only based on reasons they would endorse). However, if failing to honour their assumed wishes regarding a specific decision-making process nevertheless significantly improved one's ability to respect their autonomy in the first sense, it may be that one has done more to respect their autonomy overall.

The student essay and Editor's choice by Sara Kate Heide[15] also explores surrogate decision-making for those with diminished autonomy, including older persons with dementia. It is a beautifully written personal reflection and qualitative exploration of how seniors conceive of quality of life. In her experience working in care homes, she finds, it is often not so much about pursuing what is in their "medical best interests" that matters to seniors, but rather respecting their own sense of autonomy by helping them to maintain their lifelong sense of personal identity. In another essay, Mike King and Hazem Zohny deal with use of non-human animals in research.[16] These animals do not have decision-making autonomy in the sense that humans do, and might therefore be thought to require "paternalistic" treatment according to what is in their best interests. However, that is not the standard that is applied to non-human animals; rather, they are used instrumentally, as in lab research, and then euthanized. King and Zohny argue that, however bad this treatment is for the animals, it is also psychologically distressing to the human scientists who are charged with doing the experimentation and killing. They suggest that animal ethics committees ought to take steps to help reduce this "psychological burden" in humans.

Finally, a number of essays add to a welcome shift in focus for medical ethics, toward broader socio-structural and historical issues: Christina Richie[17] argues that pharmaceutical companies have an obligation to reduce their carbon footprint, for the sake of the environment; Pugh *et al*[18] analyse trade-offs in the use of "inaccurate" COVID-19 tests for effective public health policy at a national level; Milne *et al*[19] map out a model for participatory governance in handling of massive amounts of data in the context of large-scale biobanks; Pierre *et al*[20] share the results of their study on physician attitudes and behaviours toward incarcerated patients; and Yeo-Teh and Tang[21] address researchers' obligations to the public in conducting studies on stem-cell based therapies for autism spectrum disorder, given the ways that even poor quality research in this area is likely to be taken up by parents and other laypeople hoping for a "cure." It is heartening to see the *Journal of Medical Ethics* continue to publish essays ranging from the detailed analysis of a specific clinical case study (like Baby T) to philosophical discussions of key concepts, like autonomy, in the context of cutting-edge technological innovations (the PPP), to appraisals of systemic issues in society.[22]

REFERENCES

1. Jardas E, Wasserman D, Wendler D. Autonomy-based criticisms of the patient preference predictor. *J Med Ethics* 2022;48(5):304–10.
2. Wilkinson D, Pillay T. Surrogate decision making in crisis. *J Med Ethics* 2022;48(5):290.
3. Pillay T, Noureldein M, Kagla M. Commentary to 'surrogate decision making in crisis. *J Med Ethics* 2022;48(5):291–2.
4. Jackson B, Horsey K, Spearman A. Surrogate decision making in crisis. *J Med Ethics* 2022;48(5):297–8.
5. Kavati AB, Ramirez F. Nursing commentary to "Surrogate decision-making in crisis". *J Med Ethics* 2022;48(5):293–4.
6. Wilkinson D. Surrogate uncertainty: who decides? *J Med Ethics* 2022;48(5):295–6.
7. Gillam L. The zone of parental discretion: an ethical tool for dealing with disagreement between parents and doctors about medical treatment for a child. *Clin Ethics* 2016;11(1):1–8.
8. Alderson P. Children's consent and the zone of parental discretion. *Clin Ethics* 2017;12(2):55–62.
9. Coulson-Smith P, Fenwick A, Lucassen A. In defense of best interests: when parents and clinicians disagree. *Am J Bioeth* 2018;18(8):67–9.
10. Diekema DS. Parental refusals of medical treatment: the harm principle as threshold for state intervention. *Theor Med Bioeth* 2004;25(4):243–64.
11. Wilkinson D, Nair T. Harm isn't all you need: parental discretion and medical decisions for a child. *J Med Ethics* 2016;42(2):116–8.
12. Shalowitz DI, Garrett-Mayer E, Wendler D. The accuracy of surrogate decision makers: a systematic review. *Arch Intern Med* 2006;166(5):493–7.
13. Sharadin NP. Patient preference predictors and the problem of naked statistical evidence. *J Med Ethics* 2018;44(12):857–62.
14. John SD. Messy autonomy: commentary on patient preference predictors and the problem of naked statistical evidence. *J Med Ethics* 2018;44(12):864.
15. Heide SK. Autonomy, identity and health: defining quality of life in older age. *J Med Ethics* 2022;48(5):353–6.
16. King M, Zohny H. Animal researchers shoulder a psychological burden that animal ethics committees ought to address. *J Med Ethics* 2022;48(5):299–303.
17. Richie C. Environmental sustainability and the carbon emissions of pharmaceuticals. *J Med Ethics* 2022;48(5):334–7.
18. Pugh J, Wilkinson D, Savulescu J. Sense and sensitivity: can an inaccurate test be better than no test at all? *J Med Ethics* 2022;48(5):329–33.
19. Milne R, Sorbie A, Dixon-Woods M. What can data trusts for health research learn from participatory governance in biobanks? *J Med Ethics* 2022;48(5):323–8.
20. Pierre K, Rahmanian KP, Rooks BJ, *et al*. Self-reported physician attitudes and behaviours towards incarcerated patients. *J Med Ethics* 2022;48(5):338–42.
21. Yeo-Teh NSL, Tang BL. Moral obligations in conducting stem cell-based therapy trials for autism spectrum disorder. *J Med Ethics* 2022;48(5):343–8.
22. Blumenthal-Barby J, Boyd K, Earp BD, *et al*. Pandemic medical ethics. *J Med Ethics* 2020;46(6):353–4.