

Respectful **Debiasing** in Empirical Ethics

Shang Long Yeo

Centre for Biomedical Ethics

National University of Singapore

Bias and Empirical Ethics

- Empirical evidence about the public's judgments → normative claims about what we should do
- But some moral judgments may be biased, unreliable or untrustworthy

Bias and Empirical Ethics

Case study 1: We are deciding how to allocate scarce medical resources:

- Treatment 1 is expensive and is less cost effective, but it has identified victims (who, e.g., have been prominently featured in the news).
- Treatment 2 is cheaper and more cost-effective, but it doesn't have identified victims.

We survey the public's moral judgments. Some judge that we morally should fund Treatment 1. We worry that this is caused by a bias toward identified victims, making (some) judgments in favour of Treatment 1 unreliable (Smids et al., 2024)



Credit: Pixabay on Pexels

Bias and Empirical Ethics



Credit: Sora Shimazaki on Pexels

Case study 2: We are deciding between policies for preventing the spread of a deadly infectious disease. We find that some members of the public give inconsistent verdicts about the same choice under different frames -- e.g. when we frame the choice in terms of lives saved by each policy, they choose policy A; when we frame in terms of lives lost, they choose policy B. We worry that these moral judgments are the

- **Results of framing effects and are hence unreliable** (Tversky and Kahneman, 1981)
 - Controversial – see Daniels (2015) and Dreisbach and Guevara (2019) for contrary opinions about cases
- **Q: How should we deal with potentially biased moral judgments?**

Why should I care?

- Whether (and to what extent) biases afflict moral judgments will affect normative justification for policy
- Bias prominently discussed in psychology and economics, but less so in bioethics (Hofmann, 2023)
- Good to have a defensible method to deal with bias
- Bias in clinical settings:
 - Clinical ethics committees might be biased towards hospital interests or healthcare professionals' perspectives (Magelssen et al., 2014)
 - Doctors might be affected by status quo bias when considering withdrawing treatment (Wilkinson et al., 2019)

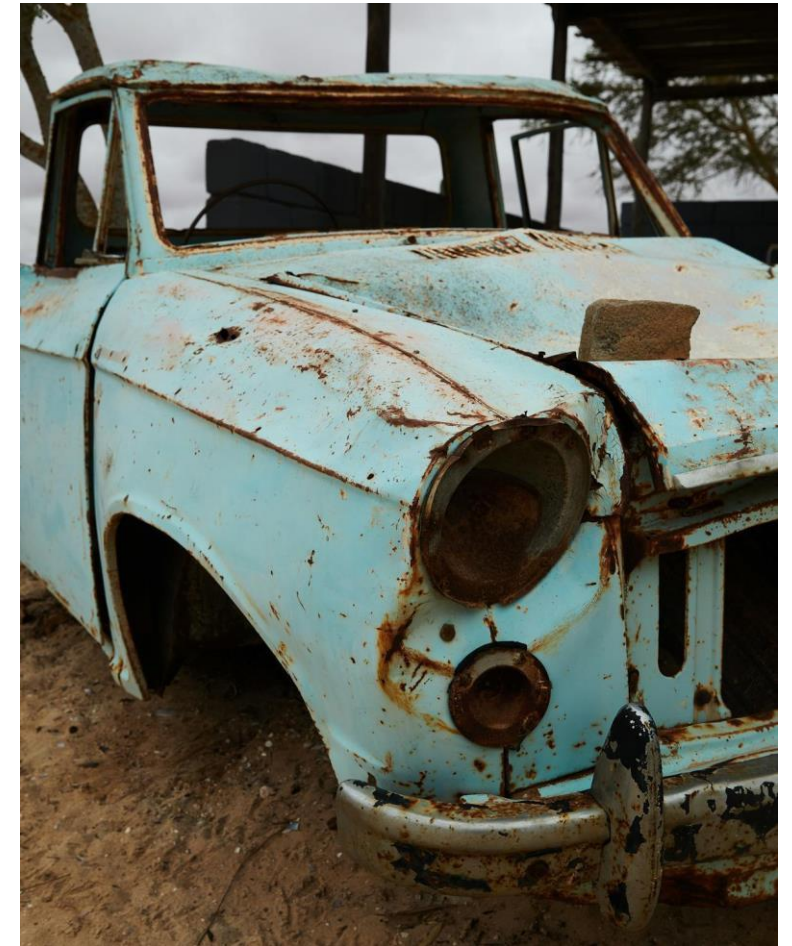
How to deal with bias/unreliability?

- Two extreme options:
 - **Evaluate and exclude:** Identify judgments that are unreliable, and exclude those from consideration.
 - **Accept as is:** Simply allow the potentially unreliable judgment to influence policy.
- Evaluate and exclude ensures that better quality moral judgments influence policymaking (Savulescu et al., 2021)
 - Compare: poorly-informed non-moral judgments
- But evaluate and exclude might suffer from moral/political issues...

Issues with Evaluate and Exclude

#1: Inappropriately evaluates people's agential capacities

- Treats agents as “things in need of repair” (Kateb, 1992, pp. 88, 230 in Carter, 2011),
 - undermining their outward dignity
- Example from Carter (2011): Professor interacting with student in the classroom vs interacting with the same student in the public sphere
- Carter: State should treat citizens as agents, by refraining from evaluating their agential capacities past a minimal threshold (opacity respect)

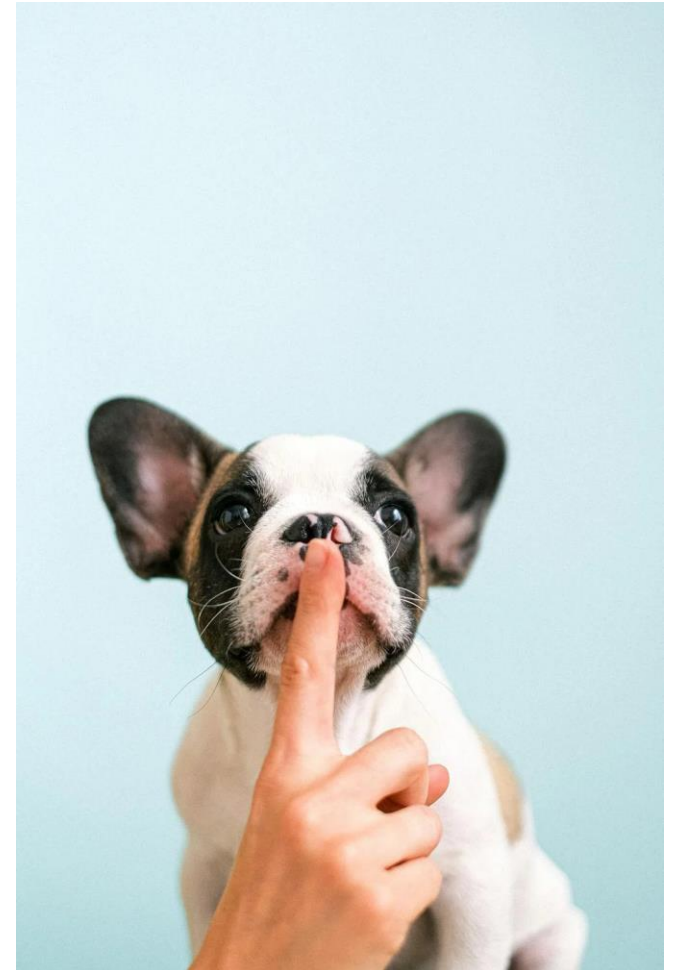


Credit: Laker on Pexels

Issues with Evaluate and Exclude

#2: Undercuts/silences people's reasons

- Prevents people from putting up reasons for consideration in the first place (Flowerree, 2023)
- To see badness, compare with how majority rule overrules a minority judgment (still exerts a vector of force).



Credit: Anna Shvets on Pexels

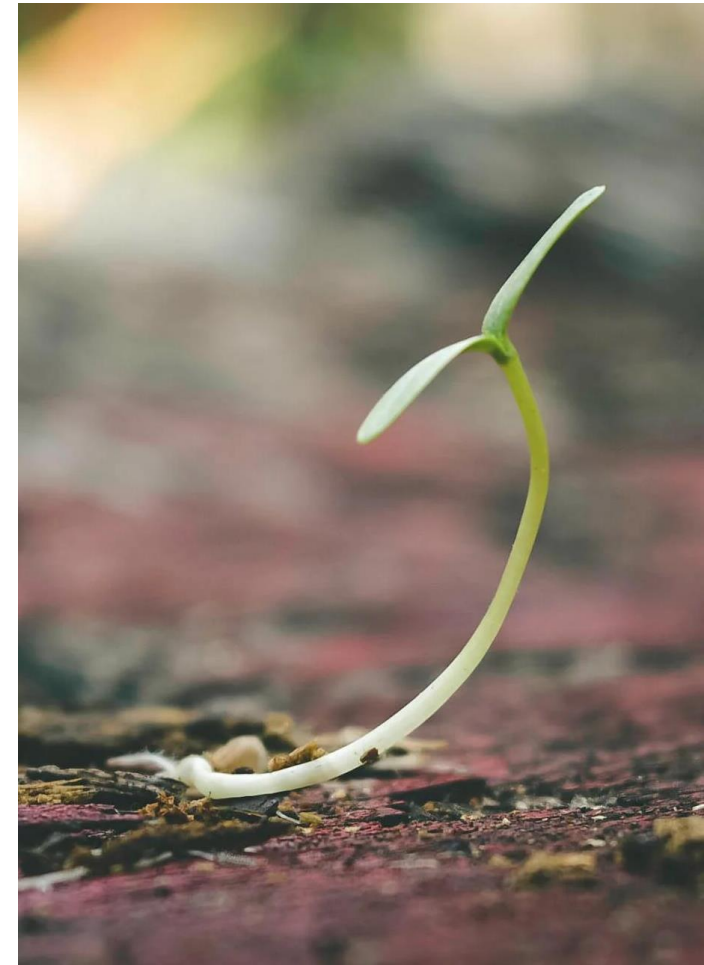
Issues with Evaluate and Exclude

- Evaluate and exclude is disrespectful + undercuts/silences their reasons
- But the other extreme (accepting judgments as is) creates unreliability
- Any other alternatives? I explore the middle ground between these two extremes
 - Outline three options
 - Evaluate their moral features

Three Respectful Debiasing Options

#1: Improve the conditions of elicitation, and accept whatever judgments people have in those conditions

- Difference: specifically choose reliable conditions
- For the identifiable victims effect: make both potential beneficiaries equally well identified (e.g. make the non-identified victim equally visible and highlight their need for help, Smids et al. 2024)
- For framing effects: put frames side by side to encourage comparison (Druckman, 2001)



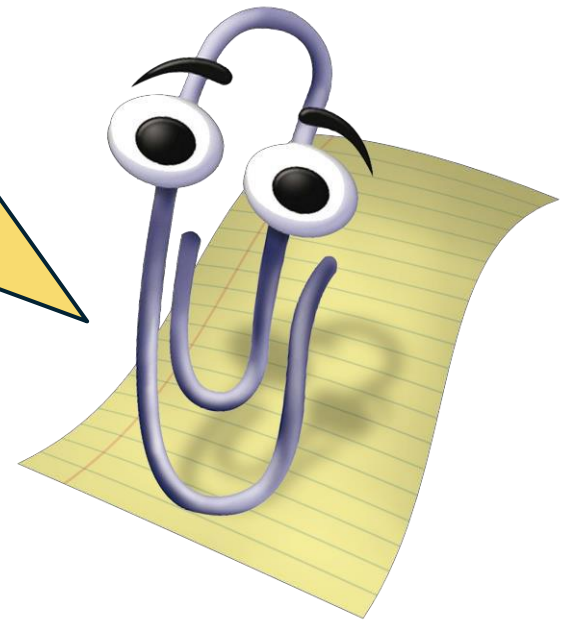
Credit: Gelgas Airlangga on Pexels

Three Respectful Debiasing Options

#2: Suggest to the person that they may be unreliable, but leave them to decide whether to change their judgment

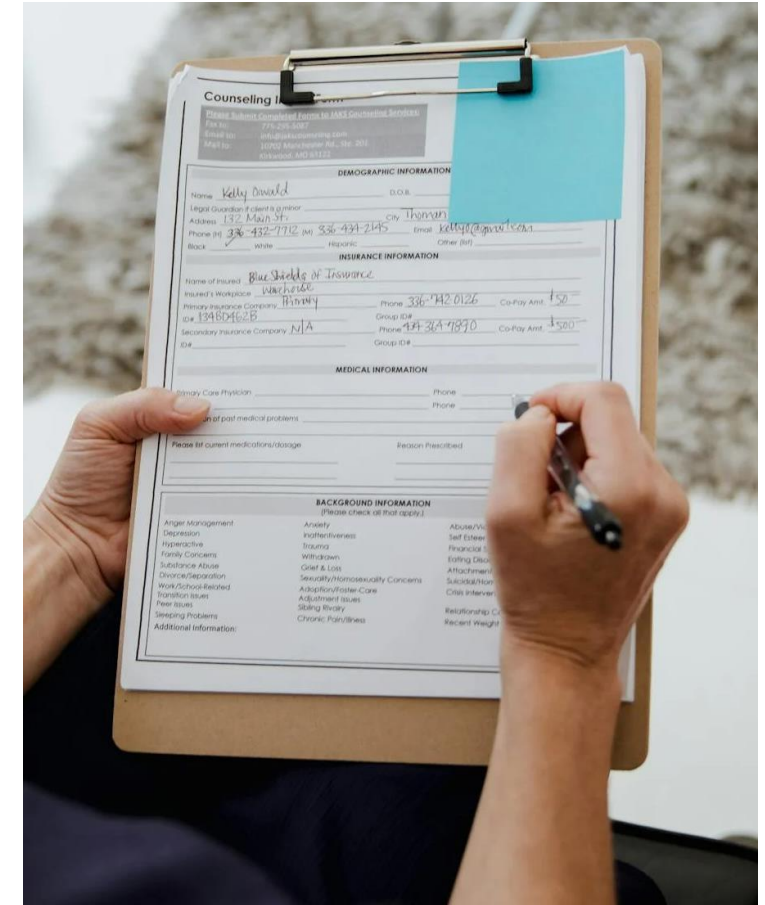
Example: “The identifiable victim effect happens when people are willing to spend much more money to save the lives of identifiable victims than to save statistical victims. Some have argued that this is a bias that makes our moral judgments unreliable. Your judgment may be caused by this identifiable victim effect -- would you like to change it?”

- Difference: merely suggests, does not impose exclusion
- Suggesting/hypothesizing is morally different from endorsing (Flowerree, 2023, p. 978)



Three Respectful Debiasing Options

- **#3: Survey the person's standard of bias, and only exclude if their judgments are biased according to a standard they endorse**
- Example: "Do you think governments should exclude a moral judgment from influencing policy if that judgment could be caused by the identifiable victims effect?"
 - If they answer 'yes' and their judgment is in favour of an identified victim, then exclude
- Difference: doesn't rely on an externally-imposed standard



Credit: RDNE Stock project on Pexels

Evaluating Respectful Debiasing

- With respect to the two moral criteria imposed on Evaluate and Exclude:
 - Whether it disrespectfully evaluates people's agential capacities
 - Whether it silences/undercuts people's reasons
- And consider other potential issues

#1: Improving conditions of elicitation

- Doesn't disrespectfully evaluate anyone's moral judgments – merely puts people in better conditions
- Doesn't silence/undercut their reasons – just accepts judgments made in better conditions
- Other potential downsides:
 - Feels somewhat paternalistic/involves interference in rational will
 - Doesn't allow for learning

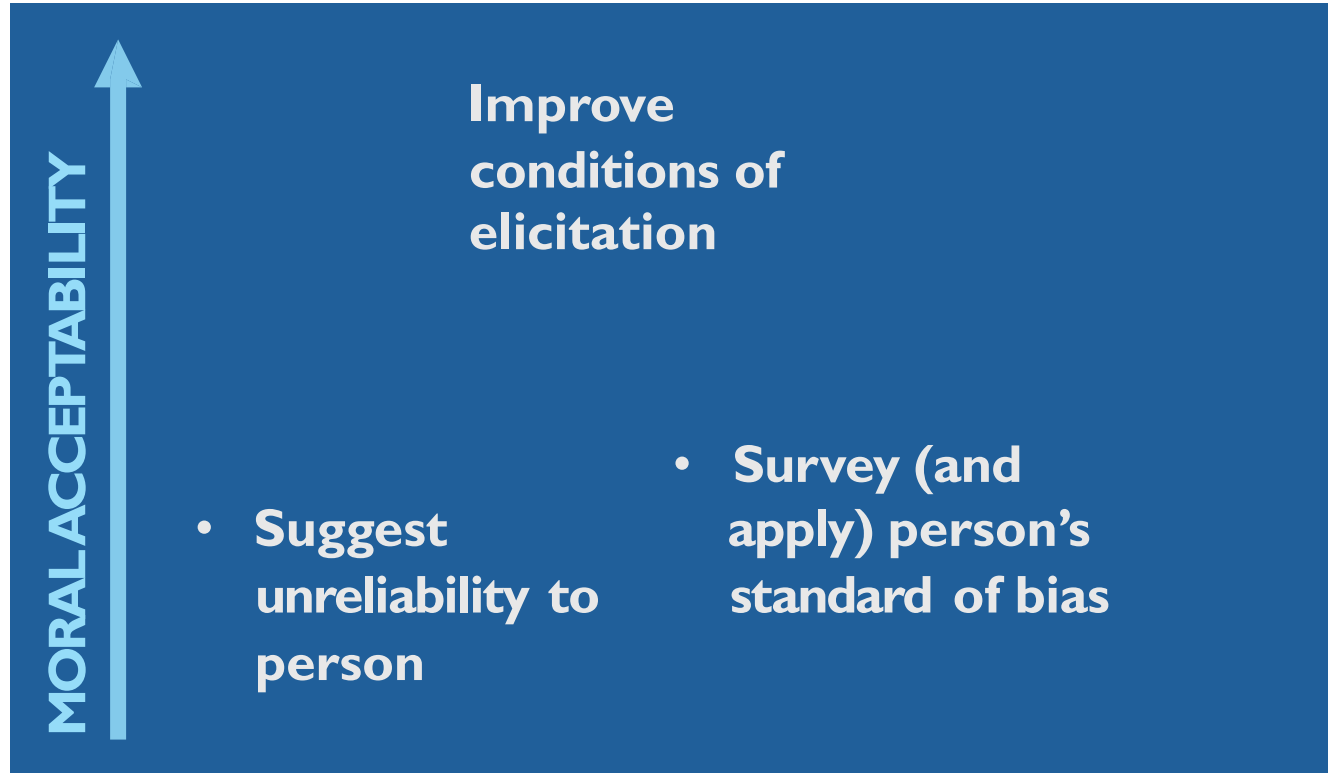
#2: Suggesting unreliability

- Disrespectfully evaluates on the way to suggesting unreliability. Perhaps less disrespectful if it's a mere suggestion?
- Doesn't silence or undercut their reasons – leaves it up to them to decide whether to change their judgment
- Other potential issues:
 - Might be gaslighting/inappropriately inducing doubt
 - But might be good for learning

#3: Surveying person's standard of bias

- Doesn't disrespectfully evaluate person – rather it's the person evaluating themselves
- Does undercut their reasons – but perhaps justifiable insofar as we hold the person to their own standards
- Other potential issues:
 - Unsure about the degree to which it induces doubt, supports learning

| | Does the method disrespectfully evaluate? | Does the method undercut or silence reasons? |
|---|--|---|
| Improve conditions of elicitation | No | No |
| Suggest unreliability to person | Yes | No |
| Survey (and apply) person's standard of bias | No | Yes but maybe justified |



| | Does the method disrespectfully evaluate? | Does the method undercut or silence reasons? | Further potential issues |
|---|--|---|--|
| Improve conditions of elicitation | No | No | <ul style="list-style-type: none"> • Feels somewhat paternalistic, involves interference in rational will • Doesn't allow for learning |
| Suggest unreliability to person | Yes | No | <p>Might be gaslighting/inappropriately inducing doubt</p> <ul style="list-style-type: none"> • But might be good for learning |
| Survey (and apply) person's standard of bias | No | Yes but maybe justified | <ul style="list-style-type: none"> • Unsure about the degree to which it induces doubt, supports learning |

Thanks for your
attention!

References

Carter, I. (2011). Respect and the Basis of Equality. *Ethics*, 121(3), 538–571. <https://doi.org/10.1086/658897>

Daniels, N. (2015). Can There be Moral Force to Favoring an Identified over a Statistical Life? In I. G. Cohen, N. Daniels, & N. Eyal (Eds.), *Identified versus Statistical Lives: An Interdisciplinary Perspective*. Oxford University Press.

Dreisbach, S., & Guevara, D. (2019). The Asian Disease Problem and the Ethical Implications Of Prospect Theory. *Noûs*, 53(3), 613–638. <https://doi.org/10.1111/nous.12227>

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, 22(1), 91–101. [https://doi.org/10.1016/S0167-4870\(00\)00032-5](https://doi.org/10.1016/S0167-4870(00)00032-5)

Flowerree, A. K. (2023). When to Psychologize. *Australasian Journal of Philosophy*, 101(4), 968–982. <https://doi.org/10.1080/00048402.2022.2157032>

Hofmann, B. (2023). Biases in bioethics: A narrative review. *BMC Medical Ethics*, 24(1), 17. <https://doi.org/10.1186/s12910-023-00894-0>

Kateb, G. (1992). *The Inner Ocean: Individualism and Democratic Culture*.

Magelssen, M., Pedersen, R., & Førde, R. (2014). Sources of bias in clinical ethics case deliberation. *Journal of Medical Ethics*, 40(10), 678–682. <https://doi.org/10.1136/medethics-2013-101604>

Savulescu, J., Gyngell, C., & Kahane, G. (2021). Collective Reflective Equilibrium in Practice (CREP) and controversial novel technologies. *Bioethics*, 35(7), 652–663. <https://doi.org/10.1111/bioe.12869>

Smids, J., Bomhof, C. H. C., & Bunnik, E. M. (2024). ‘Empathy counterbalancing’ to mitigate the ‘identified victim effect’? Ethical reflections on cognitive debiasing strategies to increase support for healthcare priority setting. *Journal of Medical Ethics*. <https://doi.org/10.1136/jme-2023-109646>

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458. <https://doi.org/10.1126/science.7455683>

Wilkinson, D., Butcherine, E., & Savulescu, J. (2019). Withdrawal Aversion and the Equivalence Test. *The American Journal of Bioethics: AJOB*, 19(3), 21–28. <https://doi.org/10.1080/15265161.2019.1574465>