The Anthropocentric Bias of Anthropic Reasoning: A Case of Implicit Dualism

Marc A. Burock

ABSTRACT

Methodological anthropic reasoning (MAR), popularized by Bostrom ([2002]), aims to correct for observation selection bias by appealing to observer-relative information. I show that MAR's inferential structure is not uniquely tied to observers but applies to any set of entities subject to selection uncertainty. By miscasting a general epistemic problem as uniquely anthropic, MAR obscures its metaphysical assumptions and bypasses established probabilistic methods. Once stripped of its observer-centric framing and functionally reduced, anthropic reasoning collapses into ad hoc inference—forcing a choice: either acknowledge the metaphysical specialness of observers or concede there is no reason to privilege one physical pattern over another.

1. Introduction

It's nearly impossible to begin a paper on anthropic reasoning without causing the reader to take on a definite state—either one of intuitive agreement or one poised to discredit everything that follows as pseudoscience. Within the fields of philosophy and physics, if you haven't already taken a side, it's likely because you haven't spent much time with the anthropic literature that has accumulated over the past 50 years.

The basic idea of the Anthropic Principle, first explicitly stated by Carter [1974]—though anticipated by Dicke ([1961]) and Collins and Hawking ([1973])—is that what we expect to observe in nature must be compatible with our existence as observers (Barrow and Tipler [1986]). For instance, we wouldn't expect to find ourselves in an environment with an ambient atmospheric temperature of 10,000K, as that temperature is not compatible with our biology. This so-called Weak Anthropic Principle (WAP), though logically uncontroversial, has continued to generate debate: some argue it is merely tautological or vacuous (Mosterín [2005]) while others regard it as a modest but essential constraint on cosmological reasoning (Weinstein [2005]; Barnes [2021]; Helbig [2023]).

WAP retains broad logical appeal because it can be flexibly interpreted depending on context. Yet this flexibility obscures precisely what does the epistemic work. If compatibility of background conditions with any complex structure suffices, observers lose their special role (Earman [1987]; McMullin [1993]). WAP enables a slippage between philosophically distinct categories—observers versus general life-forms—fueling confusion in anthropic debates.

While initially a concern for physicists, the anthropic principle was later refined by philosophers aiming for greater epistemological rigor (Leslie [1983], [1997]; Bostrom [2002]; Dorr and Arntzenius [2017]). In particular, Bostrom's project centered on formulating a general methodology to correct for observation selection effects—statistical bias that tends to "creep in when we evaluate evidence that has an indexical component" (Bostrom [2002], p. 57). Such evidence involves primitive observer-relative statements, where observers are entities for whom these statements matter. I call this approach methodological anthropic reasoning (MAR), a statistical framework distributing probabilities across observer populations to correct biases from our observational standpoint.

For the purposes of this paper, I do not deny some version of the weak anthropic principle, though I will not explicitly defend it either. My focus lies specifically on methodological anthropic reasoning: the attempt to treat first-person observerhood as epistemically primitive within a statistical framework, and to use this notion to correct for selection biases. I argue that this methodology, while influential, fails to justify its conceptual commitments, and that once those commitments are made explicit, its distinctive role in probabilistic inference collapses.

In the following sections, I present several well-known anthropic thought experiments from the literature. The goal is not to solve these puzzles or adjudicate between competing versions of MAR, but to adopt a meta-perspective: to examine the structural features of the puzzles, reflect on the sources of epistemic indeterminacy they rely on, and clarify the often-unstated metaphysical commitments that animate anthropic reasoning. After exposing the more general epistemic problems that give rise to anthropic puzzles, I will argue that MAR's methodology fails to connect with standard selection effects in statistics, is probabilistically and theoretically incoherent, and fails to address the functional reducibility of indexical statements—requiring either an explicit reckoning with the problem of consciousness or a full collapse into redundancy.

2. The General Sampling Assumption

We are often introduced to methodological anthropic reasoning through compelling thought experiments. Bostrom highlights a compelling story of *Emeralds*, initially attributed to John Leslie, meant to pump our intuition toward an almost undeniable conclusion through anthropic reasoning:

> Imagine an experiment planned as follows. At some point in time, three humans would each be given an emerald. Several centuries afterwards, when a completely different set of humans was alive, five thousand humans would each be given an emerald. Imagine next that you have yourself been given an emerald in the experiment. You have no knowledge, however, of whether your century is the earlier century in which just three people were to be in this situation, or in the later century in which five thousand were to be in it. . . The sensible bet, therefore, is that yours is instead the later century of the two. (Bostrom ([2002], p. 63)

This intuition—that you belong to the larger group centuries later— feels correct. If you're an arbitrary person holding an emerald, and there are two groups you could have come from, you likely came from the group with more members. The example becomes even more compelling when expanded to trillions. Thought experiments like *Emeralds* point toward one form of anthropic reasoning where the number of observers bias our credence in one situation versus another. However, the same reasoning applies to non-observer scenarios *Balls* as well:

> Suppose there are two groups of green balls. Group A contains three green balls, and the other group B contains five-thousand green balls. Suppose you have one of those green balls, but you don't know its origin. Which group do you suppose your ball came from?

This stripped-down version contains no observers, yet the intuition from *Emeralds* persists—your ball more likely came from the larger group. After all, there are many more ways in which your green ball came from the larger group. It appears that MAR applies similarly to plain-old inanimate objects, independent of any observer-relative bias. That result may seem surprising given the literature's focus on observer selection effects. While non-observer type anthropic arguments (like the possibility of carbon-atoms) are common using WAP, the anthropic methodological procedures made prominent by Bostrom and others focus on observer-related effects. Consider his so-called Self-Sampling Assumption (SSA):

> (SSA) One should reason as if one were a random sample from the set of all observers in one's reference class.

When applied to *Emeralds*, this unmodified version of SSA places a uniform distribution on all observers in the reference class (emerald receivers), making the probability of belonging to the larger group 5000/5003. A parallel principle for non-observers would be the General Sampling Assumption:

> (GSA) In the absence of a specified selection mechanism, one should reason as if an object were a random sample from the set of all objects in that object's reference class.

Applying GSA to *Balls* goes like this: I don't know which group my green ball came from, therefore I should reason as if the ball in my possession was a random sample from the collection of all green balls. I assign a uniform probability measure, and conclude that the probability my green ball came from the larger group is also 5000/5003. The structural isomorphism between *Balls* and *Emeralds*, the mathematical methodology, and the resultant calculations suggest that being an observer is not necessary for the general numerical framework that underlies SSA. After all, SSA is a form of counting and weighing a specific kind of entity (observers) and the methodological procedure of counting doesn't strongly depend upon the type of object being counted.

One core difference remains: SSA references both observers as objective entities and a particular indexical observer ("you"). It shifts between an indexical perspective and external observers, suggesting indexical information has special status. SSA has reasoners reasoning about themselves, while GSA has reasoners reasoning about external objects. To align GSA and SSA even more closely, instead of referring to objects in GSA, we could reformulate it:

> (GSA) In the absence of a specified selection mechanism, one should reason as if each observation of an object were a random sample from the set of all observations in that observation's reference class.

This reformulation closely mirrors SSA without conflating indexical and external perspectives. If we treat "an object" as "myself," SSA emerges naturally. But then, what does SSA offer that GSA doesn't? Can one accept SSA while rejecting GSA? Why should observers receive special treatment? Defining an observer's reference class is difficult, but no more so than defining a heap in Eubulides' paradox. Your indexical status—and the total number of observers—seems irrelevant to the general methodology. MAR puzzles don't require observers at all, except in the trivial sense that only meaning-capable entities can interpret propositions. I'm not concerned here with defending or rejecting GSA on metaphysical grounds. In the next section, I'll highlight the core epistemic problem that makes GSA, SSA, SIA[1] and other anthropic heuristics appear as potential solutions in the first place.

3. The Indeterminate Selection Problem

Bostrom acknowledges that methodological approaches to anthropics can apply to ordinary, third-person objects. Noting that anthropic puzzles like *Emeralds* can be recast in physicalist terms by mapping observers to material brain-states, Bostrom then assumes a ball is placed into an urn for each brain. Random selection over balls can then replace indexicals:

> But what exactly did change when we removed the indexical element? If we compare the two last examples, we see that the essential disparity is in how the random samples were produced... In the second of the two examples, there was *a physical selection mechanism* that generated the randomness... In the other example, by contrast, there was no such physical mechanism. Instead, there the randomness did somehow *arise from each observer considering herself as a random sample from the set of all observers*. But there was no physical randomization mechanism at work analogous to selecting a ball from an urn. (Bostrom [2002], p. 136)

This is clarifying. A common criticism against SSA-type reasoning is that the physical selection mechanism of observers is ill-defined, making it difficult to establish a robust empirical measure over a class of observers. Philosophers and scientists have long noted that different mechanisms of random selection led to different probability calculations, even for well-defined, non-indexical problems.

What distinguishes an anthropic problem is the absence of a physical selection mechanism to guide how we measure outcomes. But as previously demonstrated, we can construct coherent probability scenarios without indexical observers and without specified selection mechanisms. In *Balls*, I didn't explain how the ball came into your possession—perhaps you simply found it beside you upon waking. This ambiguity creates the confusion: how do we reason about uncertainty when we know potential outcomes exist and an outcome was selected, but no mechanism is offered? This is the anthropic dilemma.

Neither indexicality nor being an observer is unique to this epistemological question. However, anthropic reasoning assumes this type of uncertainty—when occurring over observers—can be clarified by adopting approaches like SSA or SIA. We must ask: if these procedures apply equally to non-anthropic situations with similar selection ambiguity, what makes them specific to anthropic reasoning? Consider another classic anthropic puzzle where our intuitions and methodologies generate conflicting credences:

God's coin toss with jackets: God flips a fair coin. If heads, he creates one person with a red jacket. If tails, he creates one person with a red jacket, and a million people with blue.

- *Darkness*: God keeps the lights in all the rooms off. You wake up in darkness and can't see your jacket. What's your credence on heads?
- *Light+Red*: God keeps the lights in all the rooms on. You wake up and see that you have a red jacket. What's your credence on heads?

Anthropic methodologies disagree on the correct credence. However, a non-indexical version captures identical uncertainty:

God's coin toss with balls. God flips a fair coin. If heads, he creates one red ball. If tails, he creates one red ball and a million blue balls.

- *Darkness*: You wake up in a dark room holding one of the balls but can't see its color. What's your credence on heads?
- *Light+Red*: You wake up in a lit room holding a red ball. What's your credence on heads?

This version strips away any connection to observers while preserving the fundamental uncertainty. The anthropic version distributes uncertainty over multiple observers, the non-anthropic version over multiple balls. Both exemplify what I call the Indeterminate Selection Problem (ISP):

(ISP) Given an outcome space where the selection process is indeterminate or unspecified, how should uncertainty (or credence) be rationally assigned across outcomes?

The principle of indifference (PI) is one answer to ISP, although things get more confusing when the puzzle suggests different ways to carve up the outcome space. Bertrand's famous chord paradox is among the notorious counterexamples to PI: what is the probability that a chord, chosen at random in a circle, is longer than the side of an inscribed equilateral triangle? Depending how random selection of chords is defined—and how PI is applied—yields different probabilities.[2] So PI is not enough to answer ISP.

Bostrom suggests that resolving ISP is central to anthropic reasoning. No mechanical device can select across multiple, separate first-person experiences, especially with observers in separated regions of a multiverse or as potential individuals not currently alive. But here lies the crux: if ISP is the primary conundrum driving methodological prescriptions in anthropics, yet ISP applies generally to all situations with indeterminate selection processes—regardless of observers—what makes anthropic reasoning unique?

I can apply GSA (of which SSA is a subset) to any non-anthropic situation involving outcomes belonging to different sized groups of similar outcomes. The methodology simply counts objects identified by reference classes—standard probabilistic reasoning. The anthropic element merely identifies the objects as observers. The subjective aspect of observers becomes irrelevant to SSA once we acknowledge ISP as the fundamental problem. The anthropic puzzles are intended to push our intuitions about one anthropic methodological technique versus another, but our intuitions are independent of observers versus marbles. Consider this shortened version of Bostrom's infamous *Presumptuous Philosopher*:

> Scientists are in a room debating between two theories about the universe:
>
> Theory A: The universe contains 1 trillion observers
> Theory B: The universe contains 1 trillion-trillion observers
>
> Both theories are equally supported by all available scientific evidence, with 50% probability to each. A philosopher interjects, saying, "According to anthropic reasoning of SIA, Theory B is one trillion times more likely. Dispute settled!"

This example was constructed to demonstrate the counter-intuitive consequence that SIA is almost 100% confident Theory B is true, overriding rigorously generated scientific credence. Yet replacing "observers" with "marbles" and adding "There is a marble in the room" to the puzzle generates identical intuitions. If scientists have solid theories about marbles, SIA-type calculations shouldn't override that evidence, marble present or not.

The methodological questions in anthropics ultimately point to ISP—a fundamental problem in probabilistic reasoning whenever selection processes are indeterminate. This suggests that anthropic reasoning substantially overlaps with general probabilistic inference. In the next section, I'll discuss the statistical literature addressing sampling bias (a field relatively overlooked in mathematical approaches to anthropics) and other general epistemic problems that MAR glosses over.

4. Length-biased Sampling and Anthropic Reasoning

The attempted methodological-mathematical correction of sampling bias, with respect to anthropic procedures using SSA or SIA, is known in statistics as length-biased sampling (Qin, [2017], p. 1). "Length" can refer to the number of objects in a group, object size, temporal duration, or number of temporal moments. Bias arises because you're often more likely to sample a length that exceeds the average from an underlying distribution. A classic example will clarify what this means:

> *The waiting-time paradox* (aka bus-waiting paradox, aka inspection paradox, aka renewal-theory paradox): At your local bus stop, the time-gap between bus arrivals is either 8 minutes or 16 minutes. These time intervals occur with equal 50-50 chance. Suppose you just arrived at the bus stop, don't see any buses, and have no other knowledge of the time or bus schedule. How long do you expect to wait until the next bus arrives?

You might reason as follows: the average time-gap between bus arrivals is 12 minutes. If I randomly 'entered' an average gap, then I am equally likely to be anywhere within that gap, so it seems reasonable that I'd expect to wait 6 minutes on average (half the average gap). But that answer is incorrect because it doesn't account for length-bias sampling. When randomly entering a gap, you're twice as likely to enter the 16-minute gap compared to the 8-minute one, requiring adjustment in your calculations.

Physicist Torahiko Terada first described a version of the *waiting-time paradox* in 1922 (Masuda and Hiraoka [2020]). This paradox might appear anthropic since it involves observers and self-sampling a particular moment in the absence of temporal self-location information, but this interpretation misses the point. The *waiting-time paradox* is observer-neutral. In computing, it becomes relevant whenever any program or process samples another temporal process to estimate inter-arrival times. No human observers are needed, and the mathematics are rigorous and well-developed, making an anthropic interpretation unnecessary metaphysical baggage.

4.1 The Mathematical Framework of Length-Biased Sampling

Rao ([1965]) first provided a statistical framework for size-biased distributions, later generalized by Patil and Rao ([1978]). An early ecological application might have been estimating wolf-pack sizes in the wild. Suppose pack-sizes range from 2-15 wolves with known frequency distributions of each pack-size, and you spot a random wolf in the distance. What's the probability distribution of the observed wolves pack size?

Doesn't this sound a lot like an anthropic puzzle? Translated to anthropics, the question becomes: what's the probability of an observer's group size, given you spotted an observer (yourself). Rao and others developed the statistical machinery to address the question of length-biased sampling, and the field has progressed significantly over the past 50 years. The math, proofs, and conjectures are well beyond this discussion, but the canonical length-bias distribution—a well-known starting point for length-bias adjustments—is relevant:

If k is the pack or general group-size and p_k is the underlying probability function of each group-size $p_k=P(X=k)$, then the length-biased distribution p_k^w is given by (Qin [2017], p. 2):

$$P(X = K|a\ group\ spotted) = \frac{w(k)p_k}{\sum_k w(k)p_k} := p_k^w$$

The length-biased distribution usually refers to the distribution where $w(k)=k$, so that the probability of each group size is weighted by its number of members or general size.[3] In anthropics, SIA uses the classic weighting $w(k)=k$, while SSA approximately corresponds to $w(k)=k_e/k$ (where k_e is the number of members in group k also in your epistemic situation). In real world applications, the appropriate weighting functions vary based on context and sampling procedures.

4.2 Conditional vs. Cross-Population Sampling

Statistically speaking, the canonical length-bias distribution should not be applied universally. Recall the *Presumptuous Philosopher*: there are two theories about the universe, one theory with a trillion observers and the other with a trillion times more, both with a prior probability of 50%. Compare this to the following ecological situation: there is either a wolf-pack of 5 wolves or of 10 wolves in the forest, but not both. The prior probability of each is 50%. Given I show you a wolf, what's the probability of the pack size associated with that wolf?

The answer remains 50% because you were assured, by stipulation in this specific puzzle, to be presented with a wolf, independent of whether the wolf-pack size was 5 or 10. There is no contingency about seeing a wolf in this set-up. The example presumes one actual world or the other obtains, but not both, and that you are presented with a wolf either way. In this case, being shown a wolf does not discriminate between pack sizes. I'll call this a *conditional sampling* problem.

Contrast this with a situation where two different-sized packs exist simultaneously in the forest, and you spot a random wolf. Here the length-bias distribution applies, as sampling occurs across multiple populations simultaneously, with selection probability proportional to population size. I'll call this *cross-population* sampling—it's well-studied, empirically validated, and analogous to the *waiting-time paradox*.

This distinction may seem subtle. In both cases, according to the puzzle statement, you were guaranteed to see a wolf with equal prior probabilities for pack sizes. Why treat them differently? Consider another example to clarify the distinction. Suppose there are two boxes. One box has one marble, and the other has one hundred. In the first experiment, a person first selects either box with a 50-50 chance, *pulls one marble out of whichever box she picks,* and shows you the marble. What box did it likely come from? As the box was chosen first, and you were guaranteed to see a marble conditional on a single box being chosen, seeing that particular marble provides no additional evidence about the marble's box. The answer remains 50-50. But, if the experimental sampling procedure differed, and the person selects a single marble from across both boxes, and then presents you that marble, it's far more likely the marble came from the hundred-marble box. This latter situation is where a length-biased distribution applies—a cross-population sampling problem.

The *Presumptuous Philosopher* represents a conditional sampling problem. Either one or the other universe *has already obtained*—determined by other factors—and you observe yourself independent of the specific universe. The actual universe was already settled, and your self-observation was guaranteed conditional on either universe, adding no discriminating evidence. It's not that SIA is wrong and SSA is right—it's that SIA, being classic length-biased sampling, shouldn't be applied to a conditional sampling problem, and SSA is irrelevant because your presence adds no additional discriminating evidence in the setup.

4.3 Different Types of Contingency

An SIA or SSA proponent might object that our approach incorrectly places the point of conditioning after universe selection rather than using one's existence to inform universe selection. They might argue that one's existence itself is evidence that should influence which universe is more likely, not just a guaranteed observation that happens after the universe is determined, but this ignores a reasonable metaphysical belief that our actual universe is settled.

So where does contingency enter anthropic arguments if your observation, as an observer, is typically guaranteed within the setup of the puzzle? There are two key sources. First, metaphysical *identity contingency*: even though your self-observation was guaranteed, anthropic reasoning assumes you could have been a different observer, despite being the observer you actually are. While identity contingency might seem plausible, it comes with a critical corollary in MAR: your self-observation must be indistinguishable from another observer's similar self-observation—this indistinguishability defines the reference class.

This creates a challenge for probability theory. If observers' experiences are truly indistinguishable, there's no meaningful sense (within Kolmogorov probability axioms) in which "being a different self-sample" constitutes a different probabilistic outcome. I'll discuss this problem later and its relation to interpretations of probability in the next section.[4]

The second source of contingency—*property contingency*—arises from differences in properties between observers within hypotheses, where properties, and not observer existence itself provides evidential weight. Recall *God's coin toss with jackets* where God flips a fair coin. If heads, he creates one person with a red jacket. If tails, he creates one person with a red jacket and a million with blue jackets. If you find yourself with a red jacket, this adds uncertainty because while the coin toss implies conditional sampling (your existence was guaranteed), having a red jacket appears contingent and not guaranteed. Well, it all depends on how you frame and interpret the problem.

If you are told at the outset "You will find yourself with a red jacket no matter what," then this is a conditional sampling problem providing no discriminating evidence either way. However, if told "You have a jacket" initially, and wake up later to find out it's red, then that's a (potentially) contingent observation providing evidence about your population. Careful problem specification is crucial, as evidential value depends precisely on what was guaranteed versus what was contingent in the setup. Property-contingency and identity-contingency easily become conflated in anthropic puzzles.

Methodological anthropic reasoning runs ahead of these deeper questions. The field has, in effect, transformed statistical indeterminacies—that apply to all objects—into seemingly profound puzzles about existence and observation, when the real meat of anthropic reasoning lies in clarifying epistemological limitations and forcing us to put our metaphysical cards on the table. In the next section, I'll examine how subjectivity undermines methodological anthropic reasoning.

5. There is No Reference Class

A common critique of anthropic reasoning centers on the reference class problem—the ambiguity in choosing which observers to count. Bostrom ([2002], p. 72) anticipates this issue, addressing borderline observers like entities with intellectual limitations, those unaware of experimental setups, those disinclined to apply SSA, and exotic observers like AIs or angels. In the end he attempts to resolve this ambiguity by focusing on subjective observer-moments rather than whole observers.

Scholars widely acknowledge that specifying the reference class represents a significant challenge. Advocates treat it as a technical issue—where the model might not capture the outcome space perfectly but closely enough for meaningful calculations (Neal [2006]; Friederich [2017]). They either defer precise definition to future discovery or assume the outcome space is close enough to do real epistemic work. Other philosophers familiar with the depth of the reference class problem remain far less optimistic that stipulating reference classes provides any practical guidance (Hájek [2007]).

My argument here is not about identifying the right reference class. Rather, I claim that the very concept of an anthropic reference class presupposes the following: that observers, as primitively first-person entities, can be treated as ontologically identifiable entities capable of being counted, sampled, or probabilistically reasoned about. I'll argue this assumption is unjustified and incoherent within MAR.

Observers enter SSA from the outset: we're told one should reason as if one were a random sample from the set of all observers in one's reference class. Lacking a selection mechanism, random sampling in SSA becomes a euphemism for applying the classical interpretation of probability to the outcome space. The classical interpretation involves partitioning possible outcomes into equipossible outcomes where no evidence favors some over others within the partition (Hájek [2007], [2023]; Gillies [2000]). This converts information about numbers of possibilities into probabilities. For finite outcomes, the probability of event X equals the fraction of possibilities in which X occurs. SSA thus presumes a classical interpretation of probability (applied to reference class sets) with Bayesian updating, coupled with equivocal commitments regarding anthropic selection indeterminacy.

5.1 What is a reference class set of observers?

Anthropic reasoning requires identifying and counting observer-moments equivalent to one's reference class. Consider an observer-moment like watching a hummingbird in a white room. While you can identify and count your own observer-moments, identifying other observer-moments equivalent to your own proves impossible. These moments exist, by definition, as first-person private experiences fundamentally inaccessible to others. This isn't external world skepticism; it simply marks a fundamental impossibility to point to or identify any first-person observer-moment in another person. If identifying another equivalent first-person experience remains impossible even in principle, the task of gathering all hypothetical equivalent observer-moments into a set becomes conceptually incoherent. You may *imagine* others having similar experiences and conceive of sets of observer-moments, but in reality, first-person experiences cannot be externally identified. Since MAR relies on constructing such equivalence sets, the ontological status of these sets demands scrutiny.

Two broad interpretations may explain anthropic sets. First, a subjectivist interpretation that treats cross observer-moments as epistemic constructs. An individual can conceive a collection of multiple first-person experiences and ground anthropic reasoning on this epistemic set. However, this set doesn't directly track anything in the actual world when restricted to first-person experiences. At best, it functions as a heuristic device for applying classical probability. Yet classical probability theory forbids this step since it requires distinguishable set elements. Consider a collection of 1000 indistinguishable marbles. If you cannot tell them apart in any principled way, then treating each as a distinct probabilistic outcome is unjustified. For probability purposes, indistinguishable elements collapse to a singleton. Likewise, if observer-moments are truly indistinguishable in their informational content, then their multiplicity carries no probabilistic weight under classical interpretations.

This would be an odd interpretation where the outcome space comprises subjective elements over which we impose classical probability (inconsistently). Standard subjective interpretations of probability still assume a shared, third-person accessible outcome space. If both the outcomes *and* the probabilities are subjective, we move even beyond radical subjective probability of De Finetti ([1992]) into an "anything goes" framework that may collapse the meaning of probability altogether.

Alternatively, a third-person interpretation might recast observer-moments in physicalist or functional terms, assuming observer-moments correspond to physically instantiated neural states, computational states, or information-processing states. A future understanding of consciousness would allow dispensing with first-person observer talk altogether. But this eliminates the original motivation for anthropic principles. Reducing observers to physical configurations renders the first-person indexical a side issue, raising questions about what work anthropic reasoning accomplishes. Furthermore, the indistinguishability problem persists. If configurations remain precisely indistinguishable, their multiplicity becomes irrelevant to standard probability interpretations.

Anthropic reasoning asks us to apply formal probabilistic reasoning to entities we cannot individuate, identify, or distinguish. This is not a reference class problem—it's an outcome space problem. An advocate seems to require a non-standard probability interpretation for these sets (and should clearly specify this), or to somehow make the reference class both first-person and third-person simultaneously, equivocating between perspectives to hide the problem. The way anthropic discourse slides between 'observers' and 'life-forms' supports the latter. But this is only the tip of the problem—it gets worse.

5.2. Where do observers enter physics?

Whatever observers are—whether defined functionally, physically, or consciously—their nature is fundamentally determined by our actual universe's parameters and causal structure. Even our ability to conceptualize observers depends entirely on our specific physics, philosophy, language, culture, technology, and historical moment of our species. We are not independent entities that happen to find ourselves in a universe; we are thoroughly embedded products of this universe's causal processes.

While all concepts are subject to this embeddedness, the ones that we use in our physical sciences derive determinacy through mathematical theories coupled to empirical observations that together reinforce and stabilize one another. Without the interplay between well-specified structure and empirical refinement to stabilize concepts, physics could not get far. Our current notion of an observer, as an entity possessing subjectivity, does not meet this criterion and perhaps never can until we resolve the mind-body problem.

Even with a vaguely defined observer, I still accept a weak form of anthropic reasoning that derives from basic logic. If a given structure (e.g. a molecule, planet, life-form) exists in a universe, then that universe must have the conditions necessary for that structure to exist. Once we determine the particular structure does in fact exist in our universe, it follows that our universe must have the conditions necessary for that structure. The deduction is universe-independent and so general that it applies to planets and observers alike, which again explains why cosmological anthropics can conflate observers and galaxies.

That said, the typical use of evidence in physical science works differently. For example, when we look at measurements like the cosmic microwave background (CMB) or galactic redshifts, we gain epistemically potent information that connects directly to underlying physics, thereby allowing us to distinguish between theories. These measurements carry clear discriminatory weight; different theoretical assumptions about universal parameters predict different detailed outcomes. By contrast, "I observe myself" lacks discriminatory power: this is conditional sampling and any universe we consider must already be compatible with our existence. Thus, "I observe" functions as a precondition of inquiry, of observation itself, and not a distinguishing datum.

Advocates of methodological anthropic arguments attempt to go beyond logical compatibility and preconditions, presuming that fundamental parameters might specify not just the logical possibility of observers, but also their probabilistic distribution and properties (recall the *Presumptuous Philosopher* puzzle). First, this overlooks that observers are extremely far downstream of fundamental physics. The causal path from fundamental parameters to humans involves countless stages of stellar and planetary evolution, abiogenesis, billions of years of natural selection and accidental events, and so on. The notion that one could systematically link fundamental cosmological parameters to specific observer-distribution predictions (beyond the coarse requirement of stellar stability or other proxies) is tenuous.

But most problematic is the fundamental paradox at the heart of this approach: we would need a precise structural definition of an observer to enable our fundamental theories to predict observer distributions, yet self-reflecting entities remain outside our current scientific theoretical frameworks. This creates an impossible situation. Beginning any such prediction requires either reducing the first-person observer to a third-person entity for use in a respectable physical theory—thereby abandoning the anthropic project's original premise—or undertaking a paradigm shift to include primitive subjectivity in our theoretical foundations, permitting predictions.

Even if we could reduce the observer and solve the prediction problem, there’s no reason why our solution wouldn’t apply to other complex systems, like ants. Ants also have certain distributions of colony sizes and lifespans, all arising from equally far downstream causal chains and contingent events. If different cosmological theories predicted different ant colony size distributions, we could use those differences to discriminate between theories. Indeed, using ant-thropics would be scientifically more appropriate as ants are well-defined structural systems. What then is the special role of observer-moment distributions in MAR?

A typical response to the ant-thropics challenge is, "But ants don't observe!" This reply sidesteps the central question: what does it mean to observe? If the answer invokes properties such as consciousness, self-awareness, or indexical self-reference, then the anthropic framework inherits the burden of explaining how these subjective features can be integrated into physical theory in a way that allows them to participate in probabilistic discrimination. That is precisely the concern raised in the present critique.

Scientific theories, even in their broadest physical form, lack the model-theoretic[5] scope to represent first-person observer states. A subjective conscious observer cannot be described as a variable, parameter, or structure within standard physical theories. There are no established intertheory relations or laws that bridge the language of physical theory to that of first-person phenomenology.[6] As a result, such theories are incapable of generating predictions about observers in the subjective sense.

This limitation undermines the foundation of anthropic reasoning. Empirical Bayesian inference requires that the entities involved—such as a theory T and a piece of evidence like "I observe X"—be representable within a common probabilistic space.[7] For P("I observe X" | T) to be meaningful, the statement "I observe X" must be formally definable within the model of T. In the absence of such definability, the expression is not merely low in probability—it is undefined and incoherent, like P($my\ nostalgia\ | G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}$).[8]

An alternative is to reduce the observer to a third-person physical or functional structure—or "proxy." But in that case, physical theory can already account for such systems, making anthropic reasoning unnecessary. Ant reasoning works just as well. Thus, MAR rests on a category error: it either appeals to entities that lie outside the scope of physical theory, or it repackages standard predictions into philosophically loaded language, transforming ordinary causal inference into a problem it then claims to solve.

Treating oneself as a random sample among observers—the core of SSA—does not resolve the foundational issues it purports to address. SSA operates more as a metaphysical directive than an empirical heuristic, its rhetorical force deriving from the belief that one can conceptually disconnect from one's own universe long enough to gain a God's-eye view over all possible universes. But this maneuver is illusory. One cannot transcend the causal structure of the universe in which one is embedded. Statistical random sampling requires independence; sampling oneself is arguably the most dependent sample one could take. In the next section, I will suggest that methodological anthropics can no longer assume the specialness of indexical information.

6. Lost Robots

Methodological anthropic reasoning, like the Self-Sampling Assumption, assumes that indexical statements—such as “I observe a white room”—convey information that is not reducible to third-person or non-indexical facts. The philosophical significance of indexicals gained prominence in the 1960s with the work of Prior ([1968]) and Castañeda ([1967], [1968]), and was further developed by Perry ([1977], [1979]) and Lewis ([1979]). This line of thought continues in more recent work (Perry [2021], [2022]; de Ponte [2022]). Perry characterizes the ‘essential indexical’ as follows:

> Sometimes when we use an indexical to refer to some object it conveys information about that object that other ways of referring to it would not. The indexical is essential (or at any rate very useful) for conveying that Information. (Perry [2022], p. 7)

Others more directly assert that indexical information cannot be reduced to third-person, non-indexical facts for intentional action (Babb [2016])—that it represents something over and above standard descriptive content. Anthropic reasoning often relies on this irreducibility as a foundational premise—indeed, if observers were fully reducible to non-indexical facts, the entire framework would lose its footing. Bostrom appears to take the essential nature of indexical information as self-evident: we use indexical statements constantly, and they clearly convey information. But whether such information resists reduction remains a contentious issue. He offers the following example, adapted from Castañeda:

> We can imagine (changing the example slightly) that two amnesiacs are lost in the library on the first and second floor respectively. From reading the books they have learned precisely which possible world is actual—in particular they know that two amnesiacs are lost in the Stanford library. Nonetheless, when one of the amnesiacs sees a map of the library saying “You are here” with an arrow pointing to the second floor, he learns something new despite already knowing all non-indexical facts. (Bostrom [2002], p. 133)

Few would deny that seeing the map leads to an informational update. Bostrom takes this as evidence that indexicals convey a special kind of observer information. But does the example support that claim? Some philosophers argue that the so-called essential indexical is a myth—or at least that indexicality is inessential or reducible to other facts (Millikan [1990]; Devitt [2013]; Cappelen and Dever [2013]; Magidor [2015]). If so, then anthropic reasoning need not appeal to primitive first-person perspectives. References to observers may simply reflect an efficient folk-psychological shorthand for informational states grounded in physical, third-person processes. Consider the following case of *Lost Robots*:

> The MIT Robotics Lab is testing two humanoid robots within today's technological capabilities—they self-navigate environments and run sophisticated software. Robot A and Robot B are placed on different floors of the library while powered off, then reactivated. Using onboard sensors and memory, each determines it is in an unfamiliar library. Robot A encounters a library map, labeled "You are here." It binds the indexical "You" to its internal self-model, updates its location variable <Robot_A_Location> = Map_Point_X, stores the map, and successfully navigates out of the library.

From an information-processing perspective, indexical statements like "You are here" are entirely reducible to variable substitutions within a system's internal self-model. When Robot A processes the map, it binds "You" to its self-model and "here" to specific coordinates—yielding an internal update like <Robot_A_Location> = Map_Point_X. No special first-person quality is required. The robot maintains persistent variables tracking its actuators, position relative to objects, and other sensory inputs. "Self" is nothing more than a stable set of internal variables continuously updated by sensory feedback and other states, used to guide interaction with the environment. The system continually recalibrates its place within its world model—not through subjectivity, but through standard computational feedback.

While humans may indeed experience additional phenomenological qualities or semantic meaning when processing indexicals, the functional role in updating information and behavioral changes can be fully implemented in computational systems without metaphysical extras. In computational systems, indexicals function as context-dependent variables whose values depend on the agent and situation. When Robot A or Robot B uses "I," it refers to its self-model—no metaphysical mystery required, just local variable binding.

In anthropic puzzles—such as *God's coin toss with colored jackets*—the use of indexicals requires only two capabilities: distinguishing observational states (red, blue, or indeterminate) and differentiating self from non-self. Both are handled by contemporary computational systems, entirely explicable in third-person terms. Thus, the supposed need to privilege indexical observers appears unfounded unless one assumes that observers possess metaphysical properties beyond their functional architecture. Since anthropic principles like SSA and SIA involve only Bayesian updates over informational states, it's unclear what explanatory role first-person subjectivity plays. While subjectivity may matter in metaphysics or quantum theory, it appears functionally irrelevant to MAR.

If advocates of anthropic reasoning accept that indexical information is functionally reducible to third-person facts, they face a dilemma. Either they embrace some metaphysical specialness about human observers (perhaps invoking the 'hard' problem of consciousness) or they accept that the reference class of observers expands dramatically. That class would then include countless non-conscious systems with sensors, such as basic AIs or even car navigation systems, as well as most biological organisms. Once observation is defined functionally—as the capacity to update a self-referential informational state based on sensory input—the reference class extends far beyond any tractable or principled boundary.

But the deeper problem goes beyond how large or indeterminate the reference class becomes. It's that once all relevant processing is third-person describable, the need for an observer-relative methodology disappears. Systems that update internal models based on external stimuli (whether humans, animals, or robots) can be modeled directly using standard causal and probabilistic tools. No anthropic procedure is needed. MAR becomes a solution in search of a problem.

Defenders might respond by trying to restrict the reference class to reasoning entities, or those who understand anthropic arguments, or who possess some special epistemic status. But this simply pushes the problem back a level. What counts as reasoning? Why should understanding be treated as a primitive? And what justifies treating these capacities as immune to third-person reduction? Each added criterion inserts another vague or subjective layer into a framework that purports to account for selection bias in physical theories. Yet the original goal of methodological anthropic reasoning was to clarify inference and correct for observational bias in a way that supports scientific prediction. Introducing greater conceptual indeterminacy undermines that goal, making the method not more rigorous, but more obscure. And the problem of inter-theoretic representation between observers and physical theory worsens: P(*my understanding of my nostalgia* $|G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}$) does not clarify.

The challenge remains: if indexical information is functionally reducible to computational or physical mechanisms, then there is no principled reason to treat any particular configuration as epistemically privileged in probability calculations. The original motivation for MAR rested on the supposed irreducibility of indexicals. But once those are understood as functionally reducible variables within third-person systems, the foundation collapses. The reference class in not vague; it's just not there.

At minimum, every anthropic puzzle could be rerun with the observer replaced by a sensory-enabled multi-modal LLM. The logic of the puzzle remains intact; the conclusions don't change. Even if one believes current LLMs are conscious, their self-referential reasoning is ultimately reducible to patterns in hardware and code. The anthropic puzzles survive—but the anthropic framework no longer applies.

7. Conclusions

I'm perhaps presumptuous to believe these arguments will redirect anthropic reasoning toward more clarifying ends. But here is one final way to visualize what has been going on. Advocates of anthropic reasoning reasonably imagine observers scattered throughout the universe—physical entities, understood in third-person terms, amenable to scientific explanation. Yet within each of these third-person objects, they posit something more: a first-person perspective, an indexical informational state that supposedly escapes reduction. This inner content is assumed to carry evidential weight—to inform our theories in ways that cannot be accounted for by third-person science alone. And yet, this information cannot be independently verified, individuated, or structurally defined within the very theories it is meant to constrain.

Methodological anthropic reasoning requires observer-information to be irreducibly first-person to justify its epistemic specialness, yet simultaneously demands that observers be reducible to third-person, countable entities to perform statistical calculations. If MAR insists on irreducible subjectivity, then counting observers or using third-person proxies is illegitimate; if MAR insists on counting and proxies, it must abandon irreducible subjectivity.

Advocates of MAR selectively draw upon the rigor of Bayesian and statistical reasoning yet deploy these tools in an overly flexible manner, glossing over critical intertheoretic challenges. They emphasize first-person observerhood to justify anthropic reasoning, yet quietly shift to counting third-person proxies, such as human beings or stable galaxies, when calculations demand concreteness. Similarly, they downplay the reference-class problem as merely technical, rather than recognizing it as symptomatic of MAR's underlying conceptual instability. They also fail to engage adequately with substantial critiques of indexical irreducibility that directly threaten the coherence of the anthropic project. Ultimately, MAR positions itself as a method capable of contributing empirical insights, while in practice remaining fundamentally metaphysical—masking deep philosophical assumptions behind a veneer of scientific precision.

Notes

[1] (SIA) Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist. (Bostrom and Ćirković [2003])

[2] Jaynes ([1973]) does give a specific answer to Bertrand's chord paradox by appealing to transformation group analysis, however, the specific ways of empirically drawing chords (e.g. throwing darts versus dropping straws) will yield different empirical distributions.

[3] For the waiting-time paradox, k represents the temporal length of an interarrival time.

[4] This issue is distinct from the metaphysical problem of the identity of indiscernibles (Hawley [2009]; Wörner [2021]). I do not claim that epistemically indistinguishable observers are ontologically identical. Rather, for probabilistic reasoning, *indistinguishability within the outcome*

space collapses the epistemic role of multiplicity: we do not treat indistinguishable elements as adding probabilistic weight merely by being numerous.

[5] The issue can be framed in terms of model-theoretic scope (Button and Walsh [2018]). A physical theory is formulated in a formal language that specifies a set of entities, relations, and allowable interpretations. First-person statements—such as "I observe X" or "I exist"—are not representable within this formal system, either syntactically or semantically. Without a translation function or bridging theory between the language of physics and the language of subjective experience, conditional probability expressions like P("I observe X" | T) are simply not well-formed.

[6] Intertheorectic reductions, and bridging even well-established scientific theories, remain contentious (Palacios [2023], [2024]). The hard problem of consciousness is not yet solved (Chalmers [2018]).

[7] Norton ([2010]) also critiques Bayesian reasoning in anthropics on other foundational grounds

[8] Cian and Arntzenius ([2017]) present an SSA variant defined by the credence expressions $\mathrm{C}(I\ am\ F|H) = \hat{\mathrm{C}}(\langle F\ observer\text{:}\ observer\rangle|H)$. Like others, they mix indexicals, observers, intelligent life, and populations; and inherit the problems of MAR.

References

Babb, M. [2016] 'The Essential Indexicality of Intentional Action', *The Philosophical Quarterly*, 66(264), pp. 439–457.

Barnes, L.A. [2021] 'The Fine-Tuning of The Universe For Life', in Knox, E. and Wilson, A., *The Routledge Companion to Philosophy of Physics*. 1st edn. New York: Routledge, pp. 719–730.

Barrow, J.D. and Tipler, F.J. [1986] *The anthropic cosmological principle*. Oxford [Oxfordshire], New York: Oxford University Press.

Bostrom, N. [2002] *Anthropic bias: observation selection effects in science and philosophy*. New York: Routledge (Studies in philosophy).

Bostrom, N. and Ćirković, M.M. [2003] 'The Doomsday Argument and the Self-Indication Assumption: Reply to Olum', *The Philosophical Quarterly (1950-)*, 53(210), pp. 83–91.

Button, T. and Walsh, S. [2018] *Philosophy and model theory*. First edition. Oxford: Oxford University Press.

Cappelen, H. and Dever, J. [2013] *The inessential indexical: on the philosophical insignificance of perspective and the first person*. First edition. Oxford, United Kingdom: Oxford University Press (Context and content).

Carter, B. [1974] 'Large Number Coincidences and the Anthropic Principle in Cosmology', in M.S. Longair (ed.) *Confrontation of Cosmological Theories with Observational Data*. Dordrecht: Springer Netherlands, pp. 291–298.

Carter, B. [1983] 'The anthropic principle and its implications for biological evolution', *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 310(1512), pp. 347–363.

Castañeda, H.-N. [1967] 'Indicators and Quasi-Indicators', *American Philosophical Quarterly*, 4(2), pp. 85–100.

Castañeda, H.-N. [1968] 'On the Logic of Attributions of Self-Knowledge to Others', *The Journal of Philosophy*, 65(15), pp. 439–456.

Chalmers, D.J. [2018] 'The meta-problem of consciousness', *Journal of Consciousness Studies*, 25(9–10), pp. 6–61.

Collins, C.B. and Hawking, S.W. [1973] 'Why is the Universe Isotropic?', *The Astrophysical Journal*, 180, p. 317.

De Finetti, B. [1992] 'Foresight: Its Logical Laws, Its Subjective Sources', in S. Kotz and N.L. Johnson (eds) *Breakthroughs in Statistics*. New York, NY: Springer New York (Springer Series in Statistics), pp. 134–174.

De Ponte, M. [2022] 'Demystifying the myth. Perry: Revisiting the essential indexical', *Belgrade Philosophical Annual*, 35(2), pp. 107–127.

Devitt, M. [2013] 'The Myth of the Problematic De Se', in N. Feit and C. Alessandro (eds) *Attitudes De Se: Linguistics, Epistemology, Metaphysics*. Stanford, CA: CSLI Publications (208).

Dicke, R.H. [1961] 'Dirac's Cosmology and Mach's Principle', *Nature*, 192(4801), pp. 440–441.

Dorr, C. and Arntzenius, F. [2017] 'Self-Locating Priors and Cosmological Measures', in J.D. Barrow et al. (eds) *The Philosophy of Cosmology*. Cambridge: Cambridge University Press, pp. 396–428.

Earman, J. [1987] 'The Sap Also Rises: A Critical Examination of the Anthropic Principle', *American Philosophical Quarterly*, 24(4), pp. 307–317.

Friederich, S. [2017] 'Resolving the observer reference class problem in cosmology', *Physical Review D*, 95(12), p. 123520.

Gillies, D. [2000] *Philosophical theories of probability*. London ; New York: Routledge (Philosophical issues in science).

Hájek, A. [2007] 'The reference class problem is your problem too', *Synthese*, 156(3), pp. 563–585.

Hájek, A. [2023] 'Interpretations of Probability', in E.N. Zalta and U. Nodelman (eds) *The Stanford Encyclopedia of Philosophy*. Winter 2023 Edition, https://plato.stanford.edu/archives/win2023/entries/probability-interpret.

Hawley, K. [2009] 'Identity and Indiscernibility', *Mind*, 118(469), pp. 101–119.

Helbig, P. [2023] 'Life, the Multiverse, and Fine-Tuning: Fact, Fiction, and Misconceptions', *Foundations of Physics*, 53(6), p. 93.

Jaynes, E.T. [1973] 'The well-posed problem', *Foundations of Physics*, 3(4), pp. 477–492.

Leslie, J. [1983] 'Observership in Cosmology: the Anthropic Principle', *Mind*, XCII(368), pp. 573–579.

Leslie, J. [1986] 'Anthropic Explanations in Cosmology', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, One(1), pp. 87–95.

Leslie, J. [1997] 'The Anthropic Principle Today.', in R.F. Hassing (ed.) *Final Causality in Nature and Human Affairs*. Washington, D.C.: Catholic University of America Press, pp. 163–87.

Lewis, D. [1979] 'Attitudes De Dicto and De Se', *The Philosophical Review*, 88(4), p. 513.

Liao, S. [2012] 'What are centered worlds?', *The Philosophical Quarterly*, 62(247), pp. 294–316.

Magidor, O. [2015] 'The Myth of the De Se', *Philosophical Perspectives*, 29(1), pp. 249–283.

Masuda, N. and Hiraoka, T. [2020] 'Waiting-Time Paradox in 1922', *Northeast Journal of Complex Systems*, 2(1).

McMullin, E. [1993] 'Indifference principle and anthropic principle in cosmology', *Studies in History and Philosophy of Science Part A*, 24(3), pp. 359–389.

Millikan, R.G. [1990] 'The Myth of the Essential Indexical', *Noûs*, 24(5), pp. 723–734.

Mosterín, J. [2005] 'Anthropic Explanations in Cosmology', in L.M. Valdés Villanueva, P. Hájek, and D. Westerståhl (eds) *Logic, methodology, and philosophy of science: proceedings of the Twelfth international congress*. London: King's College Publications.

Neal, R.M. [2006] 'Puzzles of Anthropic Reasoning Resolved Using Full Non-indexical Conditioning'. arXiv, https://doi.org/10.48550/arXiv.math/0608592.

Nesteruk, A. [2013] 'A "Participatory universe" of J. A. Wheeler as an intentional correlate of embodied subjects and an example of purposiveness in physics', *Journal of Siberian Federal*

University - Humanities and Social Sciences, 3(6), pp. 415–437.

Norton, J.D. [2010] 'Cosmic Confusions: Not Supporting versus Supporting Not', *Philosophy of Science*, 77(4), pp. 501–523.

Olum, K.D. [2002] 'The Doomsday Argument and the Number of Possible Observers', *The Philosophical Quarterly*, 52(207), pp. 164–184.

Palacios, P. [2023] 'Intertheoretic Reduction in Physics Beyond the Nagelian Model', in C. Soto (ed.) *Current Debates in Philosophy of Science*. Cham: Springer International Publishing (Synthese Library), pp. 201–225.

Palacios, P. [2024] 'Intertheory Relations in Physics', in E.N. Zalta and U. Nodelman (eds) *The Stanford Encyclopedia of Philosophy*. Spring 2024 Edition, https://plato.stanford.edu/archives/spr2024/entries/physics-interrelate/

Patil, G.P. and Rao, C.R. [1978] 'Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families', *Biometrics*, 34(2), p. 179.

Perry, J. [1977] 'Frege on Demonstratives', *The Philosophical Review*, 86(4), pp. 474–497.

Perry, J. [1979] 'The Problem of the Essential Indexical', *Noûs*, 13(1), pp. 3–21.

Perry, J. [2021] *Revisiting the Essential Indexical*. CSLI Publications/Center for the Study of Language & Information.

Perry, J. [2022] 'Are there essential indexicals?', *Belgrade Philosophical Annual*, 35(2), pp. 7–12.

Prior, A.N. [1968] 'Now', *Noûs*, 2(2), pp. 101–119.

Qin, J. [2017] *Biased Sampling, Over-identified Parameter Problems and Beyond*. Singapore: Springer (ICSA Book Series in Statistics).

Rao, C.R. [1965] 'On Discrete Distributions Arising out of Methods of Ascertainment', *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, 27(2/4), pp. 311–324.

Shaw, J.R. [2019] 'De Se Exceptionalism and Frege Puzzles', *Ergo, an Open Access Journal of Philosophy*, 6(20201214).

Weinstein, S. [2006] 'Anthropic reasoning in multiverse cosmology and string theory', *Classical and Quantum Gravity*, 23(12), pp. 4231–4236.

Wheeler, J.A. [1994] *At home in the universe*. New York: American Institute of Physics.

Wörner, D. [2021] 'On making a difference: towards a minimally non-trivial version of the identity of indiscernibles', *Philosophical Studies*, 178(12), pp. 4261–4278.