

Moral Twin Earth and the Problem with Reductive Naturalism

Palavi Chander

Doctoral Candidate, Philosophy, University of Delhi

Abstract

The Moral Twin Earth argument, developed by Terence Horgan and Mark Timmons (1991), poses a notable challenge to reductive naturalism in metaethics. It imagines a hypothetical world closely resembling our own, where inhabitants use moral terms such as ‘good’ and ‘right’. However, on Twin Earth, these terms are tied to different natural properties than on Earth. For instance, while Earthlings might define “good” in terms of promoting happiness (a consequentialist perspective), Twin Earthlings might associate “good” with following universal moral rules (a deontological perspective).

This raises important questions about how moral language works, particularly its meaning and reference. Suppose moral terms can point to different natural properties in other contexts, as the Moral Twin Earth argument suggests. In that case, this challenges the core claim of reductive naturalism; that moral properties are fixed and directly tied to specific natural properties. Instead of offering a universal foundation for moral language, reductive naturalism risks leading to semantic relativism, where moral discourse becomes contextual.

Ultimately, this complicates the idea that moral facts can be fully reduced to natural facts. The Moral Twin Earth argument, based on Hilary Putnam’s (1973) Twin Earth thought experiment, highlights this complexity of moral discourse and casts doubt on the feasibility of reductive approaches within moral realism.

Keywords: Hilary Putnam, Horgan and Timmons, Moral Twin Earth (MTE), reductive naturalism, Twin Earth.

Moral Twin Earth and the Problem with Reductive Naturalism

According to the Stanford Encyclopedia of Philosophy (2014), the term moral naturalism can be understood in two ways: broadly, it can refer to any approach to ethics compatible with a naturalistic metaphysical worldview (philosophical naturalism), which suggests that all facts, including moral facts, are a part of the natural world and can be understood through scientific inquiry and empirical observation. However, more commonly, moral naturalism is used to describe a specific form of moral realism that holds that objective moral facts and properties exist independently of human beliefs and are grounded in the natural world.

Moral naturalism is appealing to many because it combines the strengths of both naturalism and realism to provide a way of understanding moral facts and properties that are both rooted in the natural world and objectively binding, thereby offering a realistic and empirically grounded account of morality (Stanford Encyclopedia of Philosophy, 2014).

However, this position faces significant challenges. If moral facts or truths are real, how do they fit into a

naturalistic framework? Also, our knowledge of moral facts should work like our knowledge of physical facts (Horgan & Timmons, 1991). Failing to meet these demands makes moral properties seem mysterious, raising doubts about how we know them (Horgan & Timmons, 1991).

One way to address these challenges is through reductive naturalism, a strand of moral realism that argues that moral truths can be reduced to or explained by natural facts, often through (analytic) connections between moral and non-moral terms (Horgan & Timmons, 1991). The goal of this reductive approach is to link moral facts or truths directly to natural properties, such as biological or psychological properties; for instance, moral goodness might be reduced to evolutionary fitness or happiness (Horgan & Timmons, 1991). If reductionism succeeds, it would allow moral properties to fit within a naturalistic worldview, thereby supporting both naturalism and moral realism, which holds that objective moral truths exist (Horgan & Timmons, 1991).

However, critics argue that moral naturalism does not fully address important aspects of how we engage with moral concepts. They assert that by reducing moral properties to natural facts, moral naturalism overlooks the complexity of our moral practices and of how our moral judgments can go beyond simple descriptions of the natural world (Horgan & Timmons, 1991). The critique is directed at varieties of moral realism that incorporate moral naturalism, which, in turn, entails moral reductionism.

However, this line of thought faces significant challenges, particularly from G.E. Moore's critique known as the "Open Question Argument." Moore argues that reducing moral terms to non-moral terms fails to capture the true nature of moral discourse, raising doubts about the validity of reductive naturalism in moral philosophy (Horgan & Timmons, 1991).

In response, philosophers like Hilary Putnam (1973) have introduced the Twin Earth thought experiment to illustrate the complexities of meaning and reference in moral language. Putnam's (1973) thought experiment encourages us to reconsider how we understand moral terms and their connection to the natural world, complicating the reductive approach.

Building on this, Terence Horgan and Mark Timmons's (1991) Moral Twin Earth argument extends Putnam's Twin Earth thought experiment into the realm of moral discourse. The Moral Twin Earth argument suggests that moral terms may have different meanings in different contexts, challenging the reductive naturalist's claim that moral properties can be easily identified with natural properties. This paper will explore the implications of the Moral Twin Earth argument for moral naturalism, thereby subsequently raising important considerations for reductive naturalism.

Building on this, Putnam's Twin Earth thought experiment is extended into the realm of moral discourse by Terence Horgan and Mark Timmons's (1991) Moral Twin Earth argument. It is suggested by the Moral Twin Earth argument that moral terms may have different meanings in different contexts, which challenges the claim made by reductive naturalists that moral properties can be easily identified with natural properties. The implications of the Moral Twin Earth argument for moral naturalism will be explored in this paper, thereby raising important considerations for reductive naturalism.

Hilary Putnam's Twin Earth thought experiment

Hilary Putnam (1973) is often associated with the development of semantic externalism and his critique of traditional semantic theory, which challenges the internalist view that the meaning of a term is determined solely by factors internal to the speaker (mental states or concepts) and within this framework, psychologism specifically reduces meaning to mental phenomena. Putnam's (1973) critique targets the two central assumptions of these positions:

1. Meaning as a psychological state: understanding the meaning of a word or term relates to having ‘specific mental or psychological states’, such as having certain beliefs or memories (Putnam, 1973, p. 700). This does not imply that knowing a word's meaning requires constant conscious awareness; it treats meaning as something internal to the speaker's mind (Putnam, 1973). For instance, a person understands the meaning of ‘water’ because they associate it with a particular mental image of a clear, drinkable liquid.
2. Intension determines extension: the meaning of a term (intension; its concept or definition) helps us understand what it refers to in the real world (extension; objects it applies to). This means that two terms having the same intension imply having the same extension (Putnam, 1973). This suggests that our mental concepts of a term define what it refers to. For instance, if we define ‘water’ as a clear, drinkable liquid, then anything that fits this description would be considered water.

According to Putnam (1973), these assumptions imply that meaning (intensions) are “in the head” and that the individual speaker’s internal state is sufficient to fix the term’s reference (extension) (p. 700). He presents the Twin Earth thought experiment to challenge these assumptions as they are neither jointly nor individually tenable, as meaning and reference are influenced by external factors beyond the speaker’s mental state (Putnam, 1973).

Twin Earth thought experiment: Twin Earth is a hypothetical planet identical to Earth in every way; its inhabitants also speak English and use the term “water”, but the substance they call ‘water’ is a chemically distinct liquid from Earth’s H₂O. On Twin Earth, water refers to XYZ, a liquid that looks, tastes, and behaves like water under normal conditions (like temperature and pressure) but with a different chemical composition (Putnam, 1973). Putnam (1973) imagines that when a spaceship from Earth visits Twin Earth, Earthlings initially believe that the term “water” has the same meaning on both planets. However, upon investigation, they discover that Twin Earth’s “water” refers to XYZ, not H₂O. This demonstrates that reference is determined by the actual properties of substances rather than by the speaker’s beliefs or mental states. Earthlings would correctly report that on Twin Earth, “water” means XYZ, illustrating how identical terms can refer to different substances depending on the context (Putnam, 1973).

Scenario 1: Earth and Twin Earth in the 1750s (Pre-scientific Knowledge)

Putnam (1973) thinks of two individuals, Oscar₁ on Earth and Oscar₂ on Twin Earth, having identical psychological states (psychological duplicates). Both use the term “water” in the same way based on the observable properties of water, referring to it as a clear, drinkable liquid found in lakes and seas, and neither is aware of its chemical composition (Putnam, 1973).

Oscar₁ on Earth uses “water” to refer to H₂O, while Oscar₂ on Twin Earth uses “water” to refer to XYZ (Putnam, 1973). This showcases that despite their identical mental states and concepts, the extension of “water” differs because it is fixed by the external chemical properties of the substance, not by the speaker’s beliefs or psychological states. This difference in extension occurs even when speakers are unaware of the chemical composition of a substance (water), the extension of the term remains grounded in its natural properties-H₂O vs. XYZ (Putnam, 1973). This highlights Putnam’s (1973) insight that reference is determined externally by natural properties, even when speakers are unaware of them.

Scenario 2: Earth and Twin Earth in the 1950s, Post-Scientific Discovery

By the 1950s, scientific advancements in chemistry established that water is composed of H₂O, distinguishing it from XYZ on Twin Earth, thereby clarifying the distinction between the two substances

(Putnam, 1973). However, the reference for “water” remained unchanged; it continued to refer to H₂O on Earth and XYZ on Twin Earth because the reference is tied to the natural properties of the substances, not the beliefs of individual speakers (Putnam, 1973).

Putnam (1973) notes that ordinary speakers rely on experts, like chemists, to establish references for natural kind terms like water, ensuring the term remains accurate and consistent. This division of linguistic labour presented by Putnam (1973) ensures that terms remain consistent and scientifically grounded, even if most people lack detailed knowledge of the substances they denote. For instance, “water” refers to H₂O on Earth, regardless of whether an average speaker understands its chemical structure.

This highlights Putnam’s (1973) stance that reference is shaped by social context, as ordinary speakers rely on the expertise of others, reflecting the collective knowledge, to use terms meaningfully, thereby demonstrating that reference (fixing) is a collective and socially distributed process rather than an individual cognitive endeavour. This further helps to reiterate Putnam’s (1973) point that reference for natural kind terms like “water” is determined by external, natural properties rather than by the speaker’s mental states, beliefs or naturalistic descriptions, e.g., clear, drinkable liquid.

According to Putnam (1973), natural kind terms like “water”, “gold”, and “tiger” are rigid designators in Kripke’s framework, meaning they refer to the same underlying entity (e.g., H₂O for water) in all possible worlds where that entity exists (p. 711). However, Putnam (1973) also highlights that such terms are indexical: their reference depends on the actual properties of substances in the speaker’s environment. For instance, when someone points to a glass of liquid and says, “This is water”, the term refers to the specific substance in the glass, identified by its natural composition (H₂O). Both Kripke’s and Putnam’s views emphasise that the reference of natural kind terms is fixed by the natural properties of the substance (Putnam, 1973).

Putnam’s (1973) framework shows that there is no real disagreement when speakers from Earth and Twin Earth refer to “water”, even though the term refers to different substances (H₂O on Earth and XYZ on Twin Earth) because the reference for terms like “water” is determined by the natural or intrinsic properties of the substance itself, like its chemical composition, not by mental states or linguistic form. This illustrates that differences in extension (e.g., H₂O vs. XYZ) are grounded in external natural properties and not in the meaning of the term, hence there is no room for disagreement.

Horgan and Timmons Moral Twin Earth

In Putnam’s thought experiment, Horgan and Timmons (1991) note that the term “water” on Earth rigidly designates ‘H₂O’, while on Twin Earth, “water” designates ‘XYZ’, illustrating that the two terms, despite looking identical, have different meanings due to their distinct referents. In other words, the term “water” in Twin Earth English is not translatable by our orthographically identical terms (Horgan & Timmons, 1991, p. 458). This non-translatability shows that meanings are tied to the properties of the substances they denote, not just linguistic form. This implies that terms with different referents, despite identical spelling, cannot be considered equivalent, highlighting the contextual dependence of meaning (Horgan & Timmons, 1991).

Horgan and Timmons (1991) further propose the semantic competence argument, asserting that competent speakers intuitively understand the semantic norms governing their language, which supports the idea that “water” has a specific, rigid designation based on its causal relationship with the natural property it refers to.

Horgan and Timmons (1991) extend the Twin Earth scenario to moral terms, suggesting a similar analysis could apply. They propose a hypothetical Moral Twin Earth where moral terms like “good” or “right” might refer to different properties in different contexts or societies, just as “water” does (Horgan & Timmons, 1991).

Moral Twin Earth Scenario: imagine a planet that is almost exactly like Earth in every way in terms of geography, society, culture, and even language. On this Moral Twin Earth, there are Twin Earthlings who use moral terms like “good” and “right” in ways that are very similar to how we use them on Earth for instance, they also debate moral questions, prioritize moral values in decisions, and even have terms like “good” and “bad”, are orthographically identical to ours (Horgan & Timmons, 1991).

If explorers from Earth visited Moral Twin Earth, they would likely assume that the Twin Earthlings’ terms “good” and “right” share the same meanings as their English counterparts, given the similarity in spelling and usage (Horgan & Timmons, 1991). However, this assumption overlooks a key difference: the moral terms on Moral Twin Earth are tied to different natural properties than the ones on Earth (Horgan & Timmons, 1991). In other words, the natural property that causally regulates their practices of praise and blame differs from ours:

On Earth, moral terms like “right” are connected to consequentialist properties, such as maximising happiness; for instance, Earthlings might consider an action “right” if it produces the best outcomes for the greatest number of people (Chappell, 2011).

On Moral Twin Earth, moral terms are connected to deontological properties such as adherence to rules and duties; for instance, Twin Earthlings might consider an action “right” if it follows a specific moral rule, regardless of the consequences (Chappell, 2011). Horgan and Timmons illustrate that while Earthlings and Twin Earthlings seem to use their moral language similarly to regulate practices of praise and blame, the natural properties that causally influence these practices differ (Chappell, 2011). Specifically, the underlying meanings differ: on Earth, the term ‘right’ is associated with actions that maximise happiness, reflecting a consequentialist view, whereas on Twin Earth, it refers to actions that conform to the categorical imperative, highlighting a deontological perspective (Chappell, 2011). This highlights a significant divergence in the underlying meanings and how moral terms acquire their varying references, despite their superficial similarities.

Challenges to Moral Naturalism from the Twin Earth Argument

Firstly, the problem for moral naturalism, as noted by Horgan and Timmons (1991), lies in its claim that terms like “right” and “wrong” rigidly designate the same natural properties across all contexts. However, Earthlings and Twin Earthlings use moral language similarly, but the references of the term “right” differ: on Earth, it refers to actions that maximise happiness, while on Twin Earth, it refers to actions that conform to moral rules (Horgan & Timmons, 1991). This divergence suggests that moral terms do not consistently refer to the same natural properties across different contexts; instead, their meanings vary based on the moral framework in place, posing a significant challenge to moral naturalism’s assertion of fixed reference and undermining its claim of objectivity in moral discourse (Horgan & Timmons, 1991).

Secondly, if moral terms indeed rigidly designate specific natural properties, as moral naturalism suggests, then Earthlings and Twin Earthlings would be speaking different moral languages (their moral languages respectively) for instance: On Earth, “right” equates to maximizing happiness (consequentialism) whereas, on Moral Twin Earth, “right” equates to adhering to moral rules (deontology) (Horgan & Timmons, 1991). In this scenario, any disagreement between the two groups would be merely semantic, contradicting our

intuitive sense that Earthlings and Twin Earthlings are engaging in a meaningful debate about moral truths. Instead, they would be “talking past one another” as each group’s use of “right” refers to entirely different natural properties (Chappell, 2011). Their disagreement would amount to a misunderstanding about definitions, similar to Earthlings and Twin Earthlings arguing over whether “water” refers to H₂O or XYZ in the standard Twin Earth scenario (Chappell, 2011).

However, unlike the semantic divergence of “water,” moral disagreements intuitively seem like genuine conflicts about truth, not just definitional misunderstandings.

Lastly, if Earthlings and Twin Earthlings are both correct within their moral frameworks without having a common ground for genuine moral debate as their moral terms (e.g., good or right) are referring to different natural properties, then this goes against our intuition, which suggests that moral disagreements are universal and about the same moral question (Chappell, 2011). Instead, this outcome implies that such disagreements are merely relative to each group’s framework, with no universal truth being debated (Chappell, 2011). This reliance on moral relativism undermines the very aim of moral naturalism, which seeks to establish a universal and objective basis for morality.

In contrast to the view presented above, Horgan and Timmons (1991) argue that when Earthlings and Twin Earthlings meet one another and begin discussing morality, they quickly realise that they disagree on the application of “right”. Both groups believe they are discussing the same concept when they use the term, which complicates the situation.

According to Horgan and Timmons (1991), this perceived disagreement is not merely a matter of semantics; it intuitively feels like a genuine moral conflict, for instance, an Earthling might assert, “It is good to lie if it prevents harm”, while a Twin Earthling might counter, “No, lying is never good because it violates a moral rule”.

Even if both groups recognise that the terms “good” and “right” refer to different natural properties (e.g., consequentialist vs. Kantian properties), they would not interpret their disagreement as a simple misunderstanding about definitions (Horgan & Timmons, 1991). Instead, they would experience it as a profound moral disagreement (having different beliefs about what is morally true).

Both Earthlings and Twin Earthlings perceive their debate as a quest for moral truth, not as merely “talking past one another” (Chappell, 2011). Intuitively, they are debating the same moral question: What is truly right? They are not simply misunderstanding each other or using “right” differently, but engaging in what feels like a genuine moral conflict (Horgan & Timmons, 1991).

This intuitive sense of conflict on the part of the speakers in moral discourse further complicates the stance of moral naturalism, which struggles to account for how moral terms could refer to different natural properties while still allowing for meaningful and universal disagreement, as intuition suggests otherwise (Horgan & Timmons, 1991).

Hence, Horgan and Timmons (1991) have extended Putnam’s Twin Earth scenario to explore moral semantics and challenge moral naturalism; however, they argue that such an analysis (semantic externalism) does not hold in the moral domain. Natural kind terms like “water” can vary in reference without creating disagreements; however, moral terms like “good” or “right” cannot (Horgan & Timmons, 1991). Further, if moral terms refer to different properties across contexts, it undermines the objectivity and universality of moral discourse that moral naturalism endorses.

Challenges for Reductive Naturalism from the Twin Earth Argument

The challenges posed by the Moral Twin Earth argument for moral naturalism also apply to reductive nat-

uralism, as the latter is a subset of the former and shares the same foundational assumptions. A few of these challenges are outlined below:

1. **Moral Terms Are Not Fixed:** If the meaning of “right” depends on the context or culture (e.g., consequentialist properties on Earth and deontological properties on Twin Earth), it becomes impossible to reduce “right” to a single natural property (McPherson & Plunkett, 2018). In other words, reductive naturalism cannot provide a single, clear definition of “right” as the Moral Twin Earth argument suggests that moral terms like “right” can refer to different natural properties depending on the context or practices of a society (McPherson & Plunkett, 2018).
2. **Intuitive Counterexamples:** The thought experiment is intuitive: we can easily imagine meeting individuals like Twin Earthlings who use the term “right” differently from Earthlings but still genuinely believe that they are engaging in the same moral discourse as Earthlings (McPherson & Plunkett, 2018). This suggests that reductive naturalism oversimplifies moral language by not recognising the varied meanings that moral terms can have in different contexts and cultures (McPherson & Plunkett, 2018).
3. **Moral Disagreement:** Moral disagreements such as those between Earthlings and Twin Earthlings are intuitive and meaningful, not simply arising from mistaken use of definitions. This leads to skepticism about reductive naturalism's ability to capture the depth of moral debates, as different cultures and contexts might fix moral terms to distinct natural properties (McPherson & Plunkett, 2018). Moreover, moral disagreements highlight the complexity of moral language and the challenges faced by reductive theories in addressing such diversity.
4. **Moral relativism :** The Moral Twin Earth argument shows that moral terms like “good” or “right” can refer to different natural properties in different contexts (e.g., happiness on Earth vs. rule-following on Twin Earth). This variability leads to moral relativism, where moral truths are relative to each group's framework (Chappell, 2011). As a result, groups like Earthlings and Twin Earthlings may both be correct within their systems but lack a shared standard for meaningful moral debate, undermining reductive naturalism's claim to objectivity, as it suggests that moral truths are not universally applicable (Chappell, 2011).
5. **Generalization of the Moral Twin Earth:** Horgan and Timmons suggest that for any explanation a naturalist gives about how moral terms refer to certain natural properties, it is possible to conceive a hypothetical “Twin” world where the same explanation leads to the identification of a different property (Chappell, 2011). This suggests that moral terms may not consistently refer to the same concepts across contexts, challenging the notion of stable moral references and suggesting that reductive naturalism may overlook this variability.
6. **Applying the Open Question Argument to Moral Twin Earth:** In the Moral Twin Earth scenario, according to Horgan and Timmons (1991), Earthlings might ask, “Is maximising happiness always, right?” while Twin Earthlings might ask, “Is following moral rules always, right?” Despite these questions referring to different natural properties (consequentialist vs. deontological), the questions remain open because moral terms like “right” involve deeper normative considerations (McPherson & Plunkett, 2018). This challenges reductive naturalism, which asserts that moral terms can be fully equated with natural properties. If moral naturalism were accurate, these moral questions would seem closed and resolved (Horgan and Timmons, 1991). However, the ongoing openness of these questions suggests that moral terms possess a normative significance that reductive naturalism fails to explain adequately.

Conclusion

In conclusion, the examination of moral naturalism, particularly its reductive form, reveals great philosophical difficulties. While reductive naturalism seeks to anchor moral truths in natural properties, critics argue that this approach oversimplifies the complexities of moral discourse. G.E. Moore's Open Question Argument highlights the challenge of reducing moral concepts to purely natural terms, emphasizing that such reductions fail to capture the full depth of moral reasoning. Similarly, Putnam's Twin Earth thought experiment illustrates how moral meanings shift depending on context, further complicating the reductive framework. Horgan and Timmons extend this critique with their Moral Twin Earth argument, suggesting that moral properties lack a stable reference, which casts doubt on the feasibility of reducing them to natural facts. Taken together, these arguments challenge the coherence of reductive naturalism and point toward the need for a more refined understanding of the relationship between moral language and the natural world.

References

1. Chappell, R. (2011, September 5). *Philosophy, et cetera: Moral Twin Earth*. Philosophy, et Cetera. <https://www.philosophyetc.net/2011/09/moral-twin-earth.html>
2. Horgan, T., & Timmons, M. (1991). New Wave Moral Realism Meets Moral Twin Earth. *Journal of Philosophical Research*, 447–465. https://doi.org/10.5840/jpr_1991_19
3. McPherson, T., & Plunkett, D. (Eds.). (2018). Normative Ethics and Metaethics. In *The Routledge Handbook of Metaethics* (pp. 674–686). Routledge.
4. *Moral Naturalism* (Stanford Encyclopedia of Philosophy/Spring 2014 Edition). (n.d.). Stanford Encyclopedia of Philosophy. Retrieved January 15, 2025, from <http://plato.stanford.edu/archives/spr2014/entries/naturalism-moral/>
5. Putnam, H. (1973). Meaning and Reference. *The Journal of Philosophy*, 19, 699–711. <https://doi.org/10.2307/2025079>