Ludus Moralis

Utilizing Task Gamification to Improve Moral Psychology Experiments

For moral psychologists, unraveling the intricacies of moral behavior remains a challenge, attributed to three distinct issues: conflicts between *metaethical models*, contention *over normative frameworks*, and low *ecological validity* in experimental tasks. This paper proposes a novel solution: the development and deployment of video game-like moral tasks tailored for experimentation. By integrating the real-time and interactive aspects of video games into experimental tasks, this approach increases both realism and engagement, achieving higher ecological validity than traditional methodologies. This shift facilitates a deeper investigation of metaethical and normative disagreements in scenarios that more accurately reflect real-world conditions, promising to resolve empirical gaps and refine our understanding of moral mental processes. Thus, charting a new path in moral psychology research.

1. Introduction

Moral psychology researchers are increasingly leveraging cognitive science tools to investigate the cognitive processes behind moral conduct. Studies employing fMRI to observe brain activity during moral decision-making have identified key areas involved in moral cognition but have yet to fully explain how these translate into complex behaviors outside the laboratory (1). This knowledge gap underscores a need to re-evaluate current research methodologies and explore innovative approaches—a challenge that extends beyond merely gathering data; it involves rethinking task design and stimuli to better reflect real-world moral circumstances.

Current methodologies often rely on simplified vignettes or questionnaires that fall short of capturing the complexity of moral behavior. Such approaches, while informative, grapple with both logistical and conceptual issues. Not only are these experiments plagued by small sample sizes and regional biases, but they also suffer from *poor ecological validity* (2) (3), which pertains to the potential discrepancy between experimental settings and real-world moral situations. In the context of this limitation, this paper argues for the potential of virtual task environments (VTEs) designed in a video game style to address these challenges. This solution is partially founded on the notion that games are an essential aspect of human culture (4), and partly on that reality that VTEs offer immersive, complex scenarios that facilitate the evaluation of active decision-

making and more closely resemble the real conditions under which moral behavior unfolds. These environments allow for the functional exploration of morality and its underlying cognitive processes beyond abstract constructs, providing insights into how individuals navigate moral dilemmas in dynamic and real-time contexts. Moreover, the digital nature of video games enables wider and more diverse participation, reducing biases and enhancing the representativeness of the findings. By embracing the gamification of experimental tasks, moral psychology research can gain a deeper, more applicable understanding of moral cognition and action displayed in everyday life.

To make this argument, the structure of the remaining paper is as follows. Section 2 outlines the three major conceptual issues hindering moral psychology research and describes various proposals moving in the direction of rectifying these problems. Section 3 justifies the use of desktop video games as tasks over immersive virtual reality and showcases a well-designed experiment utilizing a real-time strategy video game to highlight the potential benefits of this medium. Section 4 presents a possible moral game, points to a geometric data collection opportunity unique to this methodology, and acknowledges some of the major limitations and ethical considerations resulting from this methodological shift. Finally, section 5 briefly recaps the paper, showing how the previous sections all fit together.

2. Reported Issues in Morality Research and Proposed Solutions

Moral psychology sits at the intersection of ethics and psychology, where researchers, using a blend of empirical analysis and conceptual frameworks, strive to untangle the complexities of moral judgments and decisions**.** While pure ethics focuses on delineating right from wrong actions, moral psychology aims to understand the mental processes underlying moral conduct. To do this, researchers imbue experimental stimuli with characteristics deemed to be appropriate representations of prominent moral theories and then employ a range of methodologies to collect psychological and neurological data. Impressively, moral psychology research has avoided considerable replication issues, as the results of the textual vignettes surveying are consistent between research groups (5). However, this same group suggests the field still fails to fully describe these mental processes across diverse contexts due to disagreements about metaethical theories

(rationalism vs. sentimentalism) and normative frameworks (deontology vs. utilitarianism). Moreover, there is an ongoing concern within the field that the vignettes lack complexity and emotional intensity of real-world morally difficult situations, casting doubt on their ecological validity—the capacity of experimental results to be generalized or applied to real-world conditions (2) (3). Overcoming these challenges is imperative for the field to progress to a point where its findings are not only theoretically robust but also practically relevant.

Historically, moral psychologists have debated over two competing psychological processing theories: is morality rooted in contemplative rationality or automated intuitions? In the rational processing camp, the *ethics of justice* emphasize progressively advanced reasoning about justice, such as punishment avoidance and instantiated principles (6). This perspective was later broadened to encompass reasoning about relationships and responsibilities, known as the *ethics of care* (7). In contrast, the *sentimentalist* theory of morality underscores the role of instinctive emotional reactions. It uses thought experiments involving moral taboos, like incest or cannibalism, to demonstrate how gut feelings significantly influence moral judgments (8). Meanwhile, other works propose categorizing moral behavior into four components: sensitivity, judgment, motivation, and character (9). This enduring debate among various rational and intuitive approaches, as exemplified by the theories of these leading thinkers, continues to spark disagreements among moral psychologists, propelling the field's ongoing exploration into the dynamics of moral decision-making.

Given the current lack of consensus, contemporary researchers strive to elucidate the dynamics between these two cognitive pathways in moral decision-making. The *dual-process theory*, underpinned by fMRI studies, posits that emotional intuition often aligns with moral norms, whereas rational thought tends towards meticulous deliberation (10). This assertion has faced criticism, particularly regarding the fMRI findings' reliability and the presumed superiority of utilitarian ethics (11). From these critiques, other authors have begun to advocate for cooperative models of moral decision-making. For instance, the *dual-process reflective equilibrium* approach posits that both intuitive (type I) and deliberative (type II) mental processes encompass emotional and rational elements (12). This model sees intuitive processes as facilitating reason tracking and conflict detection,

which are then subject to rationalization by later deliberative processes. This approach represents a more nuanced perspective, offering an alternative way to resolve the ongoing metaethical debates in moral psychology.

Alongside this ongoing metaethical disagreement, there is also still a millennia-long debate concerning what normative pillars of morality are load-bearing. To address this conceptual issue, investigators have relied on the trolley problem paradigm, which examines the conflict between two of the most dominant moral frameworks; deontology and utilitarianism (13). While the trolley problem's binary nature and morally unique scenario offer clear operational advantages, these qualities also inherently produce limitations. Firstly, this paradigm is normatively binary, implicitly assuming that moral decisions can be refined down to the interplay between norms and consequences (14). Secondly, these experiments have been described as lacking ecological validity. The trolley problem has been criticized for its lack of *experimental*, *mundane*, and *psychological realism*, detracting from its adequacy by being too humorous, sensationalistic, and sensorily distinct from real-world situations, respectively (2). This paradigm also predominantly tests moral reasoning through a *construct-driven* approach, focusing on abstract principles rather than real-life applicability, limiting the results' generalizability to other aspects of functional moral performance (3). A *function-led* approach, by contrast, evaluates moral reasoning within practical, real-world scenarios, offering insights into how individuals navigate moral dilemmas outside the laboratory setting, thereby providing a broader, more applicable view of moral behavior. Acknowledging these limitations is crucial for advancing toward more nuanced and realistic normative frameworks in moral psychology.

One solution being proposed to address the issues of normative disagreement and ecological validity concerns is the Agent-Deed-Consequence Model of Moral Judgment (ADC model) (15). This model works to fill the conceptual gaps of previous research by integrating virtue ethics as a third moral consideration and focusing on mundane moral dilemmas that have been developed in both low and high-stake variations. Experiments incorporating the ADC model ask participants to observe and rate the moral acceptability of moral vignettes operationalized based on the model. Each vignette is one of eight permutations of the same scenario, achieved by varying each moral aspect over a

positive/negative axis. For example, a vignette combination of **(A+D-C+**), would contain a virtuous person (**A+**), breaking a norm (**D-**), resulting in a desirable outcome (**C+**). Thus, the ADC model puts forth a more comprehensive moral model that focuses on realistic scenarios, making it a step toward conceptual completeness. Furthermore, by investigating scenarios that closely mimic real-world low-stakes moral dilemmas, the ADC model's experiments, particularly those using virtual reality (VR) vignettes, represent significant strides toward enhancing mundane and psychological realism (14). This approach not only addresses ecological validity concerns but also marks the initial steps towards adopting a function-led methodology, emphasizing practical applicability in moral psychology research.

While the stimuli operationalized in line with this new model offer a broader and more valid framework, the ADC model experiments like many of the others presented here, rely on using vignettes as their task stimuli. These abstract hypothetical snapshots can be presented using a variety of media ranging from simple textual or pictorial formats up to 3D audio-visual VR simulations. However, they fail to test judgments and decision-making processes in real-time and in an active state, and thus still lack the ecological validity necessary to truly tease apart morality. Recognizing this consistent limitation underscores the need for innovative approaches in moral psychology research. Video game-like tasks represent a promising avenue, offering dynamic, immersive environments that could more accurately mimic the complexities of real-world moral decision-making. These tasks can engage participants more deeply, eliciting responses that possibly reflect genuine moral reasoning and emotional reactions. Transitioning to such methodologies may bridge the gap between abstract theoretical models and the tangible, nuanced nature of human morality, paving the way for richer insights into moral psychology.

3. The Justification for and Potential of Video Game-like Tasks in Research

For this paper, VR is defined as "…any computer-generated experience that induces a sense of presence…" (16). This definition succinctly presents the two concepts associated with the efficacy of VR in research, which are, technological immersion and psychological presence, and is broad enough to include the entire spectrum of virtual

experiences across all technological levels. Immersion in VR is typically segmented into three categories: non-immersive, where users interact with a virtual environment through a screen; semi-immersive, which offers a more engaging experience through larger, often multiple screens or partial head-mounted displays (HMD); and fully immersive, achieved with complete head-mounted displays that isolate the user from the physical world (17). Despite the technological differences among these levels, they all aim to create the highest possible sense of user presence—the subjective experience of being in and interacting with a virtual environment. It is this feeling of 'being there,' regardless of the immersion level, that likely defines the efficacy of a VR experience. It is also true that a game inducing a high degree of presence likely fosters a greater degree of mental engagement, a construct more relevant experimentally (18). While one can be engaged without being fully present—illustrated by phenomena like inattentional blindness while driving—presence typically enhances engagement. Therefore, this paper asserts that it is the creation of both psychological components, regardless of immersion level, that is necessary to ensure behavioral research has validity.

This belief contrasts with the dominant view in cognitive psychology, which links increased immersion to greater self-reported presence. One highly-cited meta-analysis, concludes that higher immersion moderately increases subjective user presence (19). Similarly, another experiment demonstrates that the type of immersion impacts perceived presence, by comparing user experiences between desktop and head-mounted display games, using the game Smash Hit (20). However, this consensus might be premature since these reports overlook the level of task interactivity; as the meta-analysis does not consider task interactivity, and the game Smash Hit is at best moderately interactive. This oversight is further exemplified by a study where participants in high-interactivity first-person shooter games reported similar levels of presence across different immersion levels (21). Tangentially, the popularity of multiplayer online video games and esports, which may lack physical immersion but compensate with intense mental engagement, supports this perspective. Furthermore, given their wider accessibility, video games emerge as a more viable and effective platform for behavioral research, suggesting a shift towards using them over current immersive VR systems.

A strong piece of evidence supporting this reversion back to non-impressive technology comes from a group of researchers studying expertise and complex skill learning using the real-time strategy (RTS) game StarCraft II (22). The authors enumerate several strengths of this methodology, including accurate measures of motor performance and attentional allocation, direct measures of domain performance, large datasets, and numerous variables. Their experiment successfully gathered data from 3360 participants across various skill levels, identifying twelve predictive variables of in-game expertise. They suggest that RTS tasks might be as pivotal to cognitive science as Drosophila (the common fruit fly) has been to genetic biology, as the methodologies similarly permit large-scale longitudinal experiments, with a high degree of variable manipulability and accurate measures of performance. However, these authors do not enumerate the specific game dynamics that contribute to this potential. This paper suggests it is the genre's high interactivity, real-time components, and decision-making paths that mimic traditional sports and other real-world events that make it so engaging. These game dynamics, fundamental to many video games across a variety of genres, embody essential aspects of cognitive engagement and decision-making. Transposing them onto future experimental tasks in moral psychology could potentially enhance the ecological validity and depth of behavioral research, mirroring more closely the complexities of real-world moral decision-making scenarios.

4. Ludus Moralis, Geometric Data Analysis, and Limitations Discussion

The logical next step is to outline what a successful game-like moral experiment might look like. I propose it takes the form of a longitudinal function-led moral story task based on the game “The Long Drive,” where a participant must navigate a cross-country road trip. The gameplay would primarily be a driving simulator, requiring participants to navigate through common morally latent traffic scenarios, such as traffic jams and construction zones, on a predefined course. However, along this route, morally salient interactive vignettes depicting conflicts in normative theory would be probabilistically generated based on a pre-created list, like witnessing a poor child steal food. The game should more frequently generate low-stakes vignettes than high-stakes ones to enhance realism. Each vignette could be developed as a textual/pictorial vignette and as an

interactive scenario, enabling a comparison of judgments and decisions. The decision-making tasks would incorporate aspects that make RTS games so engaging, like the downstream effect of decisions on the remainder of the trip. Altogether, this game design outlines an approach that promises to deepen our understanding of moral decision-making by enhancing task engagement and realism.

Additionally, video game-like tasks support a novel type of geometric analysis of behavior, where researchers can collect the spacetime patterns (the change in x and y coordinates of the avatar in the virtual world over time) that correspond to moral actions in the game. This can be done by drawing inspiration from ethological research—the study of naturalistic animal behavior—and advanced collection techniques from data telemetry (23) (24). This approach, akin to creating a high-resolution ethogram, which is a *chart used to keep a record of every animal behavior over a specified duration*, would focus on tracking and interpreting the trajectories of virtual avatars within the game.

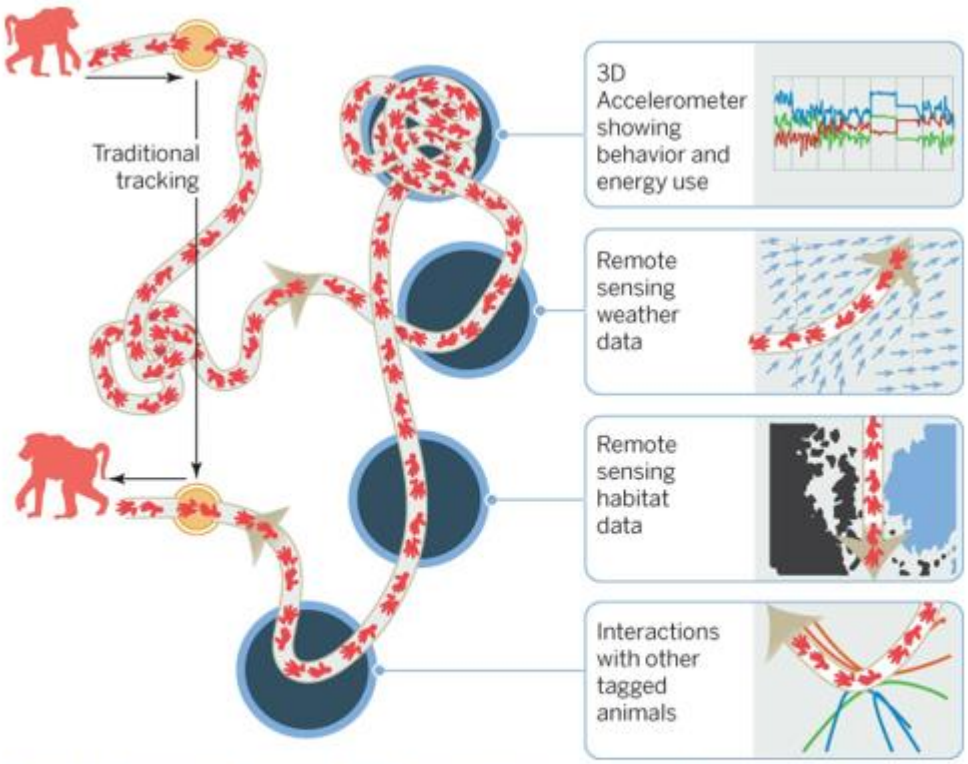

Figure 1. Geometric Analysis of Decision-Making Pathways. *Adapted from "Terrestrial animal tracking as an eye on life and planet," by R. Kays, 2015, in Science, 348(6240). DOI: 10.1126/science.aaa2478. Copyright 2015 by the American Association for the Advancement of Science. This illustration exemplifies the benefits of geometric analysis, where the pathways depicted, analogous to proposed virtual world trajectories, represent the comprehensive decision-making process within an environmental context.*

Figure 1. illustrates the potential of utilizing geometric pattern collection by showing different ways pattern analysis is enhancing how ethologists analyze animal-animal and animal-environment interactions. This same collection strategy can be applied to participant motility in VTEs. By comparing the electrical patterns of the brain to the physical ones, like keystrokes relating to in-game actions, brain regions corresponding to the processes of moral action can be identified, possibly, capturing nuances that traditional psychological metrics might overlook, particularly around the different cognitive processing theories. In sum, a strategy recognizing the importance of the journey, not just the destination, in understanding moral conduct.

In light of the positives this solution offers, there are still many methodological limitations and ethical concerns. The most significant limitation is that these experiments do not measure human responses in the 'real world'; thus, there will likely always be a separation of realism that must be considered. Secondly, it is very difficult to create morally salient audio-visual stimuli involving the thoughts and intentions of non-player virtual agents. Thirdly, if these games trigger strong moral emotions, like disgust or compassion, there is the risk of producing biased or irrational conclusions. Finally, the open-endedness that permits a meaningful spread of outcomes in RTS games also ends up greatly complicating results analysis as there are many more completion paths to consider (25). Regarding investigative ethics, firstly, the long-term effects of highly stressful virtual simulations are not well understood, suggesting that caution is required in this ethical gray area (25). Second, if the spacetimes and neurological data are to be collected and compared, extensive data collection considerations will need to be implemented to ensure standards of privacy and informed consent are met.

5. Conclusion

I acknowledge that there are additional limitations and ethical considerations not detailed in this proposal. However, the literature has clearly outlined three persistent issues obstructing the progress of moral psychology research. While various solutions have been proposed to tackle these issues individually, video game-like tasks emerge as a uniquely suited methodology to address them collectively. By leveraging highly interactive VTEs, this methodology not only enhances task ecological validity through

increased realism and engagement but also creates a battleground where metaethical and normative frameworks can be rigorously tested using stimuli that more closely resemble the real-world. The dynamic and immersive nature of video games has the potential to revolutionize our understanding of moral psychology, offering deeper and more relevant insights into human morality.

References

1. Moll J, Oliveira-Souza Rd, Zahn R, Grafman J. The Cognitive Neuroscience of Moral Emotions. In Sinnott-Armstrong W. Moral Psychology The Neuroscience of Morality: Emotion, Brain Disorders, and Development.: The MIT Press; 2007. p. 1-20.
2. Bauman CW, McGraw PA, Bartels DM, Warren C. Revisiting External Revisiting External Validity: Concerns about Trolley Problems and Other Sacrificial Dilemmas in Moral Psychology. Social and Personality Psychology Compass. 2014;: 536-554.
3. Parsons TD. Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. Frontiers in Human Neuroscience. 2015.
4. Huizinga J. Homo Ludens: A Study of the Play-Element in Culture. 1st ed. Amsterdam: Wolters-Noordhoff; 1938.
5. Skitka LJ, Conway P. Morality. In Finkel EJ, Baumeister RF. Advanced social psychology: The state of the science. New york: Oxford University Press; 2019. p. 299.
6. Kohlberg. Stages of moral development. Moral education. 1971;: 23-92.
7. Gilligan C. In a Different Voice: Psychological Theory and Women's Development Cambridge, Mass: Harvard University Press; 1982.
8. Haidt J. The moral emotions. In Davidson RJ, Scherer KR, Goldsmith HH. Handbook of affective sciences. Oxford: Oxford University Press; 2003. p. 852-870.
9. Rest J, Narvaez D, Bebeau M, Thoma S. A Neo-Kohlbergian Approach: The DIT and Schema Theory. Educational psychology review. 1999;: 291-324.

10. Greene JD, Sommerville BR, Nystrom LE, Darley JM, Cohen JD. An fMRI Investigation of Emotional Engagement in Moral Judgment. Science. 2001;: 2105-2108.

11. Sauer H. Moral judgments as educated intuitions: MIT press; 2017.

12. Cecchini D. Dual-Process Reflective Equilibrium: Rethinking the Interplay between Intuition and Reflection in Moral Reasoning. PHILOSOPHICAL EXPLORATIONS. 2021;: 295-311.

13. Thomson JJ. The Trolley Problem. The Yale Law Journal , May, 1985, Vol. 94, No. 6. 1985;: 1395-1415.

14. Cecchini D, Brantley S, Dubljević V. Moral judgment in realistic traffic scenarios: moving beyond the trolley paradigm for ethics of autonomous vehicles. AI & SOCIETY. 2023.

15. Dubljević V. Toward Implementing the ADC Model of Moral Judgment in Autonomous Vehicles. Science and Engineering Ethics. 2020;: 2461–2472.

16. Reggente N, Essoe JKY, Aghanjan Zahra M, Tavakoli AV, McGuire JF, Suthana NA, et al. Enhancing the Ecological Validity of fMRI Memory Research Using Virtual Reality. Frontiers in Neuroscience. 2018.

17. Suh A, Prophet J. The state of immersive technology research: A literature analysis. Computers in Human Behavior. 2018;: 77-90.

18. Michailidis L, Balaguer-Ballester E, He X. Flow and Immersion in Video Games: The Aftermath of a Conceptual Challenge. frontiers in Psychology. 2018;: 1-8.

19. Cummings JJ, Bailenson JN. How Immersive Is Enough? A Meta-Analysis of the Effect Immersive Technology on User Presence. Media Psychology. 2016;: 272-309.

20. Pallavicini F, Pepe A, Minissi ME. Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? Simulation & Gaming. 2019;: 136-159.

21. Yildirim C, Carroll M, Hufnal D, Johnson T, Pericles S. Video Game User Experience: To VR, or Not to VR. IEEE Games Media Entertainment (GEM). 2018.

22. Thompson JJ, Blair MR, Chen L, Henrey AJ. Video Game Telemetry as a Critical Tool in the Study of Complex Skill Learning. PLOS ONE. 2013.

23. Kays R, Crofoot MC, Jetz W, Wikelski M. Terrestrial animal tracking as an eye on life and planet. Science. 2015.

24. Kohwalter TC, Murta LGP, Esteban C. Capturing Game telemetry with Provenance. In Brazilian Symposium on Computer Games and Digital Entertainment (SBGames); 2017.

25. Pan X, Hamilton AFdC. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. British Journal of Psychology. 2018;: 395–417.