MEMORIA CIENTÍFICO-TÉCNICA PROYECTOS INDIVIDUALES
Convocatoria 2024 - «Proyectos de Generación de Conocimiento» y actuaciones para la formación de personal investigador predoctoral asociadas a dichos proyectos.

Tipos A, B y RTA

AVISO IMPORTANTE - La memoria no podrá exceder de 20 páginas. Para rellenar correctamente esta memoria, lea detenidamente las instrucciones disponibles en la web de la convocatoria. Es obligatorio rellenarla en inglés si se solicita 100.000 € o más (en costes directos).

IMPORTANT – The research proposal cannot exceed 20 pages. Instructions to fill this document are available at the website. If the project cost is equal or greater than 100.000 €, this document must be filled in English.

1. DATOS DEL PROYECTO. *PROPOSAL DATA*

IP 1 (Nombre y apellidos. *Name and surname*)**:** Ivar Rodríguez Hannikainen

TÍTULO DEL PROYECTO (ACRÓNIMO): Laboratorio sobre inferencias de estados mentales y moralidad (MSIMLab)

TITLE OF THE PROJECT (ACRONYM): Mental State Inference and Morality Lab (MSIMLab)

2. JUSTIFICACIÓN Y NOVEDAD DE LA PROPUESTA *JUSTIFICATION AND NOVELTY OF THE PROPOSAL*

2.1. Adecuación de la propuesta a las características y finalidad de la modalidad seleccionada. *Adequacy of characteristics and the purpose of selected modality.*

Moral judgments, both about what others did in the past (WP1) and what one should do in the future (WP2), depend crucially on information about our mental lives: whether the wrongdoer in question acted intentionally or merely negligently, whether others' requests for help reflect their genuine, autonomous preferences or not, and so on. This has been documented by a wealth of empirical literature from across experimental ethics and moral psychology—as we will see in greater depth in Section 2.2—and is also an explicit demand in various professional and deontic codes and even in legal doctrine.

Despite this fact, others' mental lives are inaccessible to us as moral evaluators. We cannot directly grasp whether, e.g., a defendant acted with the direct intention to harm or whether a patient is expressing a genuine or autonomous preference. Rather, these properties of third-parties' mental lives must be *inferred* in one way or another from their observable behavior, including their speech, aspects of their social identity, or even contextual cues—about where and when the behavior transpired.

Juxtaposing these two insights yields the problem that the *Mental State Inference and Morality Lab* (MSIMLab) will seek to shed light on.

(Landmark 1) Moral judgments depend crucially on information about others' mental states.
(Landmark 2) However, third-parties' mental states are not directly accessible to us.

How then is it that we (successfully or unsuccessfully) glean insights into others' mental lives in order to discern what is morally right from wrong?

The MSIMLab will aim to offer a unifying theoretical framework by which to understand how we carry out *mental state inference* in the moral domain, that explains how these processes

transpire in 'high-stakes' professional contexts (such as those that arise in medicine and law) as well as in everyday life. In short, the MSIMLab sets out to provide a contribution to basic, ***non-applied research*** within the domain of "*Mind, Language & Thought*", advancing our understanding of the cognitive processes that subserve moral reasoning, and therefore fits the characteristics of the ***non-oriented modality*** (i.e., *without* a thematic priority).

2.2. Justificación y contribución esperada del proyecto a la generación de conocimiento en la temática de la propuesta. Hipótesis de partida. *Justification and expected contribution of the project to the generation of knowledge on the theme of the proposal. Starting hypothesis*

Suppose a coworker poisons you by sprinkling white powder into your coffee. Our moral common sense dictates that whether the white powder was correctly labeled "Cyanide" or falsely labeled "Sugar" matters to their culpability. If your coworker knew that the white powder would kill you, they are extremely morally culpable. If instead your coworker reasonably believed they were sweetening your coffee as you had requested, we intuitively recognize that this action blameworthy to the same degree (Margoni & Surian, 2016; Young et al., 2007). This attenuated blame is mirrored in criminal law, where the determination of responsibility must obey *Coke's principle*–according to which an act (the *actus reus*) does not make a person guilty unless (their) mind is also guilty (the *mens rea*). This brings us to the first landmark: Our ***moral appraisals depend on information about third parties' mental states***. This applies *retrospectively*, to judgments like those of your coworker's past behavior, as well as *prospectively*, to questions about what one ought to do.

The second landmark is the orthodox view that ***third-party mental states are fundamentally inaccessible*** to us (Goldman, 2006). Rather, as social agents, we develop the ability to represent other minds throughout our lives—as abundant theoretical and empirical research on theory of mind has shown. These processes of mental state attribution, however, are plagued with difficulties: We cannot ordinarily trust people's self-report of their desires, goals or beliefs—as people might whether intentionally or unconsciously offer false reports. This raises the intriguing problem of how people ordinarily form impressions about the content of others' minds.

Combining these two insights raises the problem that the MSIMLab will attempt to solve: If our moral judgments depend on the content of others' minds, and yet we cannot grasp this content directly, how do moral agents go about establishing what others knew, wanted or intended in order to form a judgment about what is morally right and wrong?

Throughout the precursor project (the "Moral Science and Institutional Design Lab" **PID2020-119791RA-I00**; funded in the Retos de Investigación Type A modality; 9/2021 – 8/2024), my colleagues and I uncovered multiple strands of **preliminary evidence** pointing toward a role for *mental state inference* in moral decision-making beyond the determination of *mens rea*, which illustrate the need for a cohesive theory of mentalization in the moral domain.

Consider first conflicts between the moral principles of autonomy and beneficence as they transpire in the medical domain. Advance directives (ADs) provide a written record of a person's past preferences for end-of-life care. For example, an AD might state that a person would like to have treatment *withheld* under specific future conditions. Theorists and medical professionals have often noted that difficulties arise when a person's past preferences, as expressed in an AD ("T1 request"), come into conflict with their current best interests ("T2 preference"), e.g., after considerable cognitive decline.

Which element should guide treatment decisions when such conflicts arise? In a large-scale experiment (Toomey et al., 2024), we posed this exact question while orthogonally manipulating whether participants made first-person decisions (for their future self) or surrogate decisions about the treatment of close *and* distant others (i.e., for a friend or a stranger). Participants were significantly less likely to honor patients' T1 request when making surrogate decisions for strangers (~29%) and loved ones (~30%) than they were to honor their *own* T1 request in the first-person condition (~46%). Why might this discrepancy emerge between first- and third-person reasoning?

One obvious interpretation is that this discrepancy emerges from obstacles toward inferring others' authentic preferences (i.e., their mental states) that are absent when *introspecting* about one's own preferences. In two further studies, we have learned that these ***mental state inferences draw on external, observable clues***: such as the agent's identity or behavior.

Current legislation on medical aid in dying largely restricts eligibility to *adults*, whose suffering derives from a *physical disorder*, and whose prognosis is *irreversible*—and these restrictions accord with people's moral views about right and wrong (Hannikainen et al., 2024). In two nationally representative experiments, we asked why minors, and individuals suffering mental illness, or with reversible prognoses are seen as unfit to request medical aid in dying?

Using structural equation modeling, we found strong evidence that ***inferences about the patient's underlying preference explained this pattern of moral disapproval*** (see Table 1). Perceptions of stability (whether the person's preference was stable in the *past*) and satisfaction (whether the patient's preference is likely to remain in the *future*) mediated the effects of adulthood, diagnosis and prognosis. This suggests that approval of a request for euthanasia is largely explained by the degree to which the patient's observed behavior (their request) is seen as reflecting an authentic preference in this twofold sense.

Table 1. Structural equation model of eligibility criteria effects on euthanasia approval**.**

	Non-adult		**Mental disorder diagnosis**		**Reversibility prognosis**	
	%	***z***	**%**	***z***	**%**	***z***
Satisfaction	42	-7.50	38	-6.65	34	-9.79
Stability	34	-7.26	21	-4.65	22	-8.11
Reflection	15	-5.38	9	-3.90	7	-4.84
Suffering	3	-1.67$^{\#}$	6	-3.43	4	-3.50
Direct effect	7	-0.64$^{\#}$	25	-2.07*	34	-5.03
Total effect	-9.97	-7.23	-10.67	-6.70	-16.49	-12.59

Note. All p values < .001 unless otherwise noted. $^{\#}$: $p > .05$. *: $p > .01$. Satisfaction and stability jointly explain 76%, 59% and 56% of the total effects of adulthood, diagnosis and prognosis.

A similar line of reasoning can explain controversies surrounding the limits of free speech. In previous studies, we have repeatedly found that outgroup members are censored to a greater extent than ingroup members when making the same statement (Almagro et al., 2023). For example, it sounds intuitively worse for a Dutch economist to say "Productivity in Spain would greatly increase by getting rid of the *siesta*" than it does if we imagine a Spanish economist making that statement. Why might the speaker's identity have this effect? After unsuccessful attempts to explain this, our recent research found promising evidence that it may be ***driven***

by processes of mental state inference (see Fig. 1). When outgroup speakers refer to us, we are more likely to interpret their uncongenial statements as emanating from a harmful intent than when ingroup speakers make the very same statement.

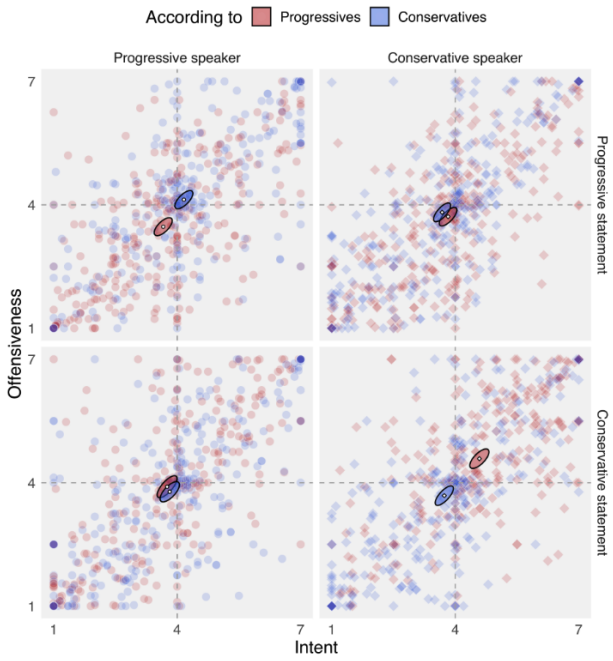


Figure 1. Inferences of harmful intent underlie patterns of offensive speech attribution.

Collectively, these three examples (of which there are likely many more) strongly suggest that moral evaluations are imbued with processes of mental state inference. The question, then, is how these processes are implemented in the mind, in light of the brute fact that we cannot directly access other people's mental lives. That is, we do not know whether the psychiatric patient request for euthanasia is any less genuine than the cancer patient's, or the Dutch economist's intent to ridicule Spain any more than it was the Spanish economist's intent.

This is the central question that the Mental State Inference and Morality Lab (MSIMLab) will pursue: How do we (accurately or inaccurately) infer others' mental lives for purposes of issuing a moral judgment?

In other areas of cognitive science, remarkable progress has been made in understanding how humans infer an agent's goal or intention from their observed behavior, by applying Bayesian reasoning principles (e.g., Baker et al., 2009; Frank & Goodman, 2012). The ***guiding hypothesis*** is that this inferential process, largely unexplored in social reasoning tasks like moral evaluation, underlies the phenomena described above. The proposed research in the MSIMLab project will address this gap by investigating *how people infer others' mental states to make moral judgments*, both about what they did in the past and what they should do in the future. In short, to decide what is right, we must distinguish offensive from inoffensive intent, authentic from inauthentic preferences, etc., and we do so by combining information about the *prior* (who the speaker is) to uncover the most likely mental state that would have led the agent to act the way they did. In this way, the MSIMLab will offer a ***novel contribution*** at the intersection of cognitive science and moral psychology, examining the applicability of Bayesian reasoning to infer mental states from behavior in processes of moral cognition.

3. OBJETIVOS, METODOLOGIA Y PLAN DE TRABAJO. *OBJETIVES, METHODOLOGY AND WORK PLAN*

3.1. Objetivos generales y específicos. *General and specific objectives*.

The ***overarching goal*** of the MSIMLab project is to leverage indirect evidence from prior research in experimental philosophy in pursuit of a unifying explanation of how mental state inference guides moral reasoning. To achieve this, the MSIMLab project will decompose this problem into three specific objectives (SOs): tackling first the question about ***mental state inference in retrospective evaluation*** **(SO1;** *how we retrospectively evaluate others*), then the question concerning ***mental state forecasting for prospective choice*** **(SO2;** *how we prospectively decide what to do*). Lastly, the project will end with the question of ***universality and variability in mental state inference*** **(SO3).** Together, WPs 1 and 2 embody the dual

temporal dimensions of Bayesian reasoning: inference from the observed past to (WP1) *explain what happened* and to (WP2) *predict what will happen* in the future. In this way, the MSIMLab project will offer a synthesis of moral cognition and its roots in mental state inference pursuing both dimensions of Bayesian reasoning and shedding light on whether the proposed mechanism emerges across cultures in WP3. This general objective will be broken down into three subgoals: How do these processes vary as a function of whether we *retrospectively* evaluate another person's behavior (WP1) or *prospectively* decide what the morally right thing to do would be (WP2)? Finally, to what extent are processes of mental state inference in moral cognition universal or variable across cultures (WP3)?

Specific Objective 1: Mental State Inference in Retrospective Evaluation

Previous research finds that our minds apply Bayesian reasoning principles to understand why agents behave the way they do (their goal). This research has fruitfully applied Bayes' Theorem to action understanding to uncover how people infer the goal of an agent (e.g., its destination) on the basis of its behavior (i.e., its motion). Baker and colleagues (2009) showed that people draw goal inferences from observed actions by inverting the process by which actions are planned in order to achieve salient or available goals (see Expression 1):

[Exp. 1] $P(Goal \mid Action, Environment) \propto P(Action \mid Goal, Environment) \times P(Goal \mid Environment)$

This kind of Bayesian inference integrates evidence from observed behavior with top-down constraints from the prior to infer an agent's objective. The formula for action understanding in Expression 1 can be adapted to the problem of mental state inference in dyadic, moral interactions. In Expression 2, the likelihood of a particular mental state given the patient's observed behavior (e.g., their speech or actions) is defined as probabilistic planning in a Markov decision process as follows:

[Exp. 2] $P(State_i \mid Behavior, Agent) \propto P(Behavior \mid State_i, Agent) \times P(State_i \mid Agent)$

Here, $P(State_i \mid Request, Patient)$ represents the probability of a specific mental state ($State_i$) given the agent's actual behavior. $P(Request \mid State_i, Patient)$ denotes the probability of observing the behavior given that the agent has that particular mental state. Finally, $P(State_i \mid Patient)$ represents the prior, unconditioned probability of the agent being in the relevant mental state—which generalizes to the set of plausible mental states for the agent.

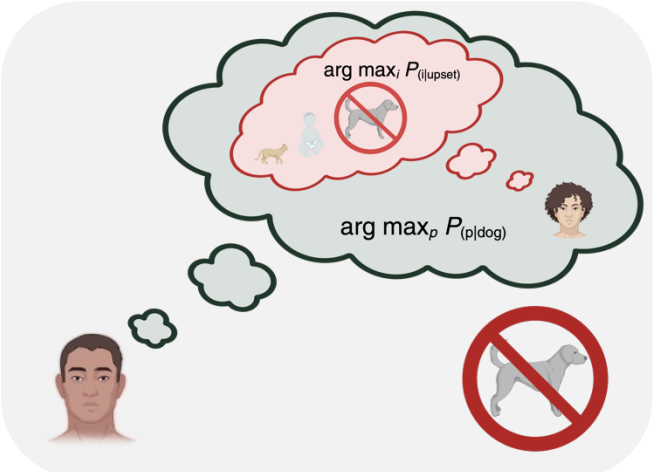


Figure 2. Bayesian mental state inference in rule interpretation.

Inspired by the evidence reviewed in Section 2.2, this inspires the hypothesis that observers might *morally* evaluate an agent's behavior by deriving the probability that it was caused by a set of plausible intentions through inverse planning: i.e., by multiplying the likelihood of making the request given a benign (/malign) intention by the prior probability of having the benign (/malign) intention.

Figure 2 illustrates this procedure in the context of rule interpretation. How do we evaluate whether a defendant's behavior violated a communal rule, e.g., that says 'No Dogs Allowed'?

The Bayesian inference model would attempt to retroactively assess whether the defendant violated the rule by inferring the rulemaker's intent from the rule's specific formulation. This prior knowledge about the rulemaker and the context (e.g., a restaurant) in which the comment was made would guide this inferential process as described in Expression 2. In Specific Objective 1, we will explore the suitability of this model across multiple contexts: inferring credibility from testimony, regret from apology, negligence from inaction, and so on.

Specific Objective 2: *Mental State Forecasting for Prospective Choice*

The second objective of the project is to understand the temporal dimension of this inference process—predicated on the idea that our behavior in social contexts often has *enduring* consequences, such that we must attempt to predict others' *future* preferences to inform our current decision about what is right or wrong. Formally, this involves defining the posterior distribution of relevant mental states (as described in Expression 2) and computing the posterior predictive distribution to simulate the target's future preferences (as shown in Figure 3; see also Pole et al., 2018).

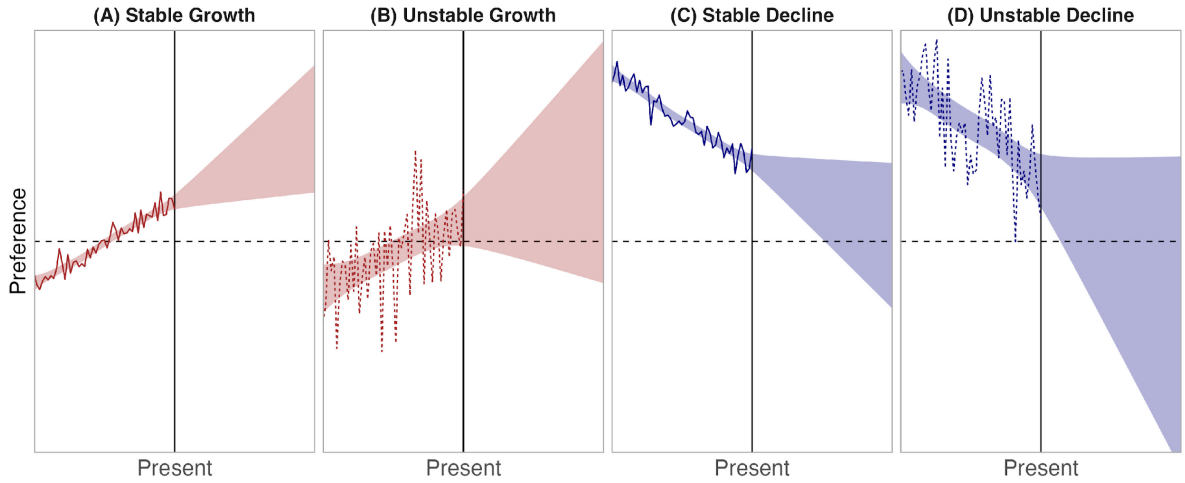


Figure 3. *Mental state forecasting of future preferences across stability and satisfaction scenarios.* Preferences are displayed as *trajectories*—illustrating how, beyond the estimation of *mean* preference strength (positions on the *y*-axis), stability (i.e., *variance* before the present) and satisfaction (i.e., *integral* after the present) intuitively influence the overall value of compliance with a third-party's preference.

For example, a doctor might use Bayesian forecasting to anticipate how a patient would (counterfactually) feel about their euthanasia request in the future—thus informing their current ethical judgment and actions. This involves using available records (up to the present), including past actions and contextual evidence, to predict future preferences. Put simply, what is the likelihood of a patient regretting their request for euthanasia based on prior preferences and medical history? By considering only past information about the patient (age, medical history, psychological state), the doctor can make an informed decision about how to proceed. For example, they might intuitively recognize that (i) *decline* scenario C is more likely to lead to regret than growth scenario A, *despite the fact* that Preference C is stronger than Preference A. Similarly, they might view (ii) instability scenarios as worse than stability scenarios, although the patient's past preference is equally strong (in C vs. D) or weak (in A vs. B).

Thus, Specific Objective 2 focuses on using Bayesian forecasting to predict future mental states (intentions, desires, beliefs) based on past behavior and asking how a forecast of potential future states guides an individual's current behavior. As new data is gathered (e.g.,

a patient's emotional state changes or new evidence about their situation arises), the Bayesian model updates its predictions to provide a more accurate forecast of future preferences. This allows for real-time adjustments in evaluating the ethical implications of one's anticipated decision.

Objective 3: Cross-cultural variability in mentalization during moral reasoning

This work package aims to explore the universality and variability of mental state inference in moral reasoning across different cultural and linguistic contexts. While much research assumes that the cognitive mechanisms underlying moral decision-making are universal, recent findings suggest that cultural norms, language structures, and social practices may shape how individuals infer and forecast others' mental states. By collecting data from diverse cultural (k = 7) settings—including the United States, Brazil, Spain, Germany, Egypt, India, and China (see Fig. 4)—we aim to test whether the processes of mentalization and their application in moral reasoning are consistent across these contexts or show systematic variation. This will allow us to engage directly with ongoing debates within experimental philosophy and cognitive science about the extent to which moral cognition is shaped by a universal cognitive architecture versus culturally specific influences.

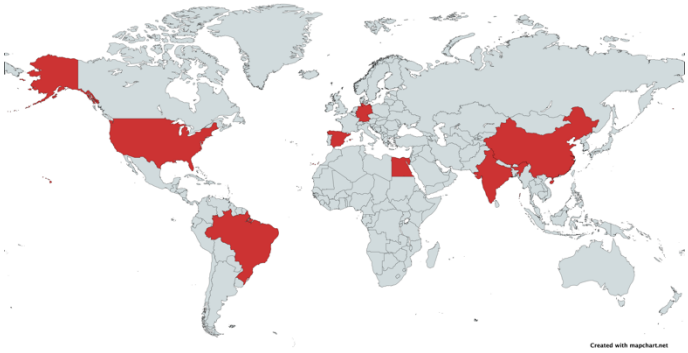


Figure 4. *Sampling plan in WP3.* Cultures are sampled from the Inglehart-Welzel cultural map to provide a stratified sample of the primary cultural regions.

To achieve this, we will replicate the experimental tasks and surveys used in WP1 and WP2, adapted for cultural and linguistic appropriateness. Our approach will integrate diverse populations and control for key socio-demographic variables, ensuring robust cross-cultural comparisons. By leveraging this global dataset, we will test hypotheses about cultural variability in retrospective evaluations (e.g., assigning blame or praise) and prospective forecasting (e.g., predicting regret or future satisfaction).

For example, we aim to investigate whether individualistic societies prioritize autonomous mental states differently from collectivistic cultures, where communal or relational aspects may play a stronger role. The findings from WP3 will help to document variation in moral cognition across cultures and thereby finesse existing theories about the extent to which processes of moral cognition are malleable by individual vs. cultural factors.

3.2. Descripción de la metodología. *Description of the methodology*.

The following methodological guidelines will govern the MSIMLab project, which draws on a suite of empirical methods to examine the retrospective uses (WP1), prospective uses (WP2), and generalizability (WP3) of mental state inference in moral cognition.

Study design. Browser-based experiments will form the basis of the MSIMLab project. These online studies will be programmed primarily in *Qualtrics*, or *jsPsych* (https://www.jspsych.org/) when the study requires the presentation of customized audiovisual and textual stimuli.

Participant recruitment. Non-probabilistic sampling in browser-based experiments will be recruited through the *Prolific* platform and the *CloudResearch* Prime Panels—two platforms with remarkable data quality. In addition, to recruit a diverse range of nationally representative samples in WP3, we will employ market research platforms such as *Netquest* (which

specializes in Spanish-speaking countries), *TGM Research* (which offers panels in Asian and African countries), *Respondi* and *Credamo*.

Sample size estimation and statistical power. Target sample sizes will be set to achieve 80% power (a β level of .20) in exploratory studies, and 95% power (a β level of .05) in confirmatory studies—both with an α level of .05. In first-stage exploratory research, effect size estimates will be drawn from published meta-analyses of the average effect (Pearson's $r = .21$) and updated subsequently with the magnitude of the observed effect in previous studies. Depending on the complexity of the target analysis, *a priori* power analyses will be derived either using closed form expressions (e.g., the *pwr* package in *R*) or empirically, by assessing statistical significance of a test result on multiple replicates of simulated data by resampling with replacement from a pilot dataset (i.e., using bootstrapping).

Pre-registration. To reduce researcher degrees of freedom and the risk of false positives, all confirmatory and replication studies will be preceded by a time-stamped pre-registration on the https://aspredicted.org/ platform detailing the study design, hypothesis, target sample size, exclusion criteria, and planned analyses—as they were conceived prior to data collection.

Statistical analysis. Study participants will often be asked to provide demographic information (like their age, political orientation, etc.) or aspects of their personality. Participants will also be presented with a study information sheet and a consent form. Once data collection is complete, anonymized data will then be analyzed collectively using statistical software, such as *R* or *Python*. Data analyses will be performed in *R* using the *rstan* and *brms* (Bayesian Regression Models using Stan) packages. For each study, we will conduct one or more statistical analyses (*t*-test, ANOVA, regression, etc.) to evaluate the focal, pre-registered hypothesis.

The risks inherent to our planned research are described in Section 3.4.

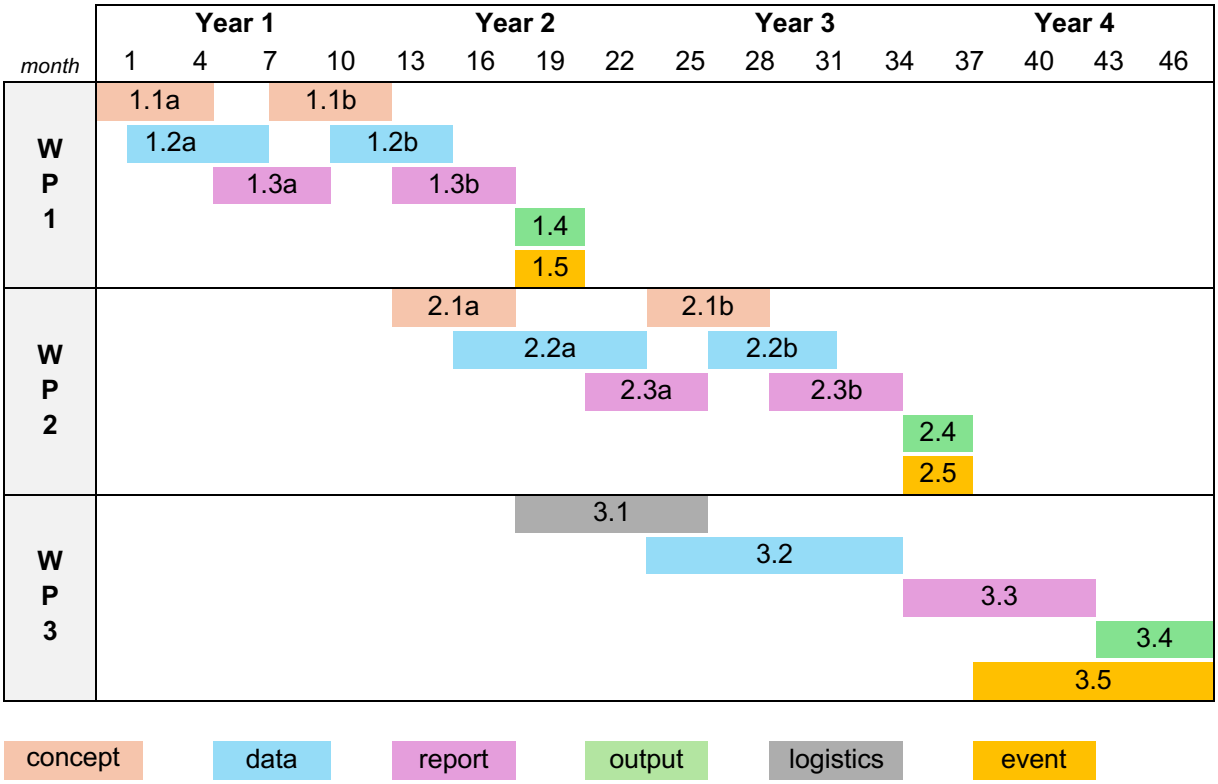
3.3. Plan de trabajo y cronograma. *Work plan and schedule*.

The project's research plan consists of three different work packages, depicted in the Gantt chart below (see Fig. 5). During WP1 (Months 1-24), we will focus on conducting the soft time pressure studies and writing a report documenting the evidence for cognitive conflict and reflection in legal interpretation. WP2 (Months 13-36) will turn to the extensions on the coordination game findings, in which we will attempt to understand why and how coordination incentives promote textualism and dampen moral reactions. Lastly, WP3 (Months 25-48) will involve re-analyzing historical survey data on moral attitudes and legislative change (e.g., abortion, end-of-life treatment, same-sex rights, etc.).

Work Package 1 involves the initial application of Bayesian inverse planning to understand the role of mental state inference in the medical and criminal domains. Task 1.1a will involve stimuli development and pre-testing of study materials. Task 1.2a involves pilot data collection, power analysis, and sample size determination. In Task 1.3a, the exploratory results will be analyzed and submitted for presentation at academic conferences. Task 1.1b will involve finalizing and running the pre-registered experiments with convenience and nationally representative samples recruited on Prolific and Netquest. In Task 1.3b the manuscripts will be updated to include the results obtained in Task 1.2b. Deliverable 1.4 will involve the submission of experimental reports on mental state inference in retrospective moral evaluation and a theoretical review piece. For WP1, the team provides specialized guidance on the role of mental state inference in retrospective culpability assessments within morality and law. **Dr. Neele Engelmann** and **Prof. Teneille Brown**, experts in law and *mens rea*, will ensure the research aligns with legal frameworks for culpability. **Prof. Kathryn Francis**, with her expertise

in moral psychology, will contribute insights of how mental state inferences shape moral evaluations. Together, they will ensure WP1 is grounded in both legal and psychological perspectives.

Figure 5. 4-Year Project Timeline in Gantt Chart Format.



Work Package 2 involves extending our research on mentalization in the moral domain by employing time-series and Bayesian forecasting techniques. The division between the tasks will be analogous to WP1: Tasks 2.1a and 2.1b concern stimuli development, task programming, and sample size determination; Tasks 2.2a and 2.2b will entail exploratory and confirmatory stages of data collection; and Task 2.3a and Task 2.3b will include the modeling, analysis, and reporting of empirical results for submission to an academic conference and a peer-reviewed journal. Deliverable 2.4 will involve the submission of a mixed-methods report combining modeling and experimental data on preference forecasting in the moral domain. For WP2, the team integrates diverse expertise of **Dr. Joanna Demaree-Cotton**, **Mr. Jorge Suárez**, **Prof. Brian D. Earp**, and **Prof. Lori Bruce** will guide the design of the studies, leveraging their knowledge of experimental philosophy and bioethics to capture the various effects influencing prospective moral judgments, with a particular emphasis on the helping professions.

Work Package 3 of the research project will involve a cross-cultural replication effort, a closing workshop and production of the primary theoretical paper. The workshop will invite a core group of collaborators who will convene to co-author the theoretical review, while also inviting submissions from academics working in related fields (such as philosophy, bioethics, cognitive science, artificial intelligence, and medicine, among others). The workshop will also serve as a platform to disseminate and discuss the results of the workshop with collaborators and colleagues. In Task 3.1, we will initiate the preliminaries of the cross-cultural replication effort including contacting international collaborators and distributing study materials. Task 3.2 will involve the implementation and execution of the cross-cultural replication study. Task 3.3 will involve integrating coauthors' feedback and writing the empirical report. Deliverable 3.4 will

involve authorship and submission of the manuscript to a theoretical outlet (e.g., *Trends in Cognitive Sciences*), and Deliverable 3.5 will involve holding the workshop on Mentalization in the Moral Domain at the University of Granada. This workshop will be geared toward producing a theoretical review paper (aimed at *Behavioral and Brain Sciences* or *Trends in Cognitive Sciences*) that critically examines the broader hypothesis about the role of mental state inference in morality from multiple angles. For **WP3**, the team combines complementary expertise to address the challenges of cross-cultural research. **Prof. Edmond Awad** and **Prof. Jim Everett**, both experts in cross-cultural methodologies, will provide strategic guidance, and **Mr. Daniel Martín**, a doctoral researcher with a specialization in moral psychology, will lead the execution of WP3, developing study materials, overseeing cross-cultural data collection, and performing the analyses.

3.4. Identificación de puntos críticos y plan de contingencia. *Identification of critical points and contingency plan*.

The MSIMLab project has identified three critical stages that could affect the viability of the proposed work plan: ***translation*** of study materials (in WP3), multi-country participant ***recruitment*** (in WP3), and ***data loss*** (affecting all WPs: WP1–3). To mitigate these risks, we have established specific strategies. For translation, we will utilize professional translation services and, when necessary, supplement with AI tools to expedite the process without compromising quality. For recruitment challenges, we will collaborate with established research companies, such as Netquest, which have a proven track record of delivering data punctually. To safeguard against data loss, robust backup protocols will be in place, with regular storage of materials and data in a secure cloud environment. Table 2 outlines the degree to which these three primary risks impact each WP, the mitigation strategies employed, and the contingency plans devised to address potential challenges.

Table 2. *Summary of Risks and Mitigation Strategies.*

	Risk	WP	Mitigation Strategies (*M*s) and Contingency Plans (*C*s)	
1.	***Translation of study materials***	1 2 3	(***M1***) Use of professional translation services to expedite the stimulus preparation phase.	(***C1***) Use of artificial intelligence tools to complement professional translation.
2.	***Multi-country recruitment***	1 2 3	(***M2***) Partnership with research companies (e.g., *Netquest*) that have previously delivered data punctually.	(***C2***) Reducing the number of data collection sites and requesting a project extension.
3.	***Data loss***	1	(***M3***) The team will regularly back-up materials and data to a 2TB cloud to monitor progress and guard against data loss.	

Additionally, I bring substantial experience in executing cross-cultural research, having led three large-scale studies (Hannikainen et al., 2019 2021 2022) involving thousands of participants across multiple international field sites. The project is further supported by an international network of ten collaborators with diverse expertise, significantly reducing risks of delay or incompletion.

3.5. Resultados previos del equipo en la temática de la propuesta. *Previous results of the team in the theme of the proposal.*

The research team brings together a group of internationally recognized experts across disciplines that are central to the MSIMLab project. Prof. Lori Bruce (Yale University) and Prof.

Teneille Brown (University of Utah) contribute extensive knowledge in ***medical ethics*** and ***psychology of law***, respectively. Dr. Joanna Demaree-Cotton (University of Oxford), Mr. Daniel Martín (University of Granada), and Mr. Jorge Suárez (University of Granada) bring specialized expertise in ***experimental philosophy***, particularly in the cognitive foundations of moral and legal reasoning. The team also includes several prominent researchers in ***moral psychology***: Dr. Neele Engelmann (Max Planck Institute of Human Development), Prof. Kathryn Francis (Keele University), Prof. Brian D. Earp (National University of Singapore), and Prof. Jim A. C. Everett (University of Kent). Prof. Edmond Awad (University of Exeter) enhances the team's multidisciplinary skills with his computational background, and unique expertise as lead author of the highly-cited 2018 *Nature* paper titled 'The moral machine experiment' (https://www.nature.com/articles/s41586-018-0637-6).

Together, this team of collaborators ensures a multidisciplinary approach to the MSIMLab project's aims, leveraging their collective track record of impactful research to guarantee high-quality outcomes. Below is an abridged list of the research team's recent publications, highlighting numerous collaborations between team members and high-impact publications that attest to the promise and likely impact of the proposed research.

2024

1. **Hannikainen, I. R.**, **Suárez, J.**, Espericueta, L., Menéndez-Ferreras, M., & Rodríguez-Arias, D. (2024). Legal provisions on medical aid in dying encode moral intuition. *Proceedings of the National Academy of Sciences*, 119(44). https://doi.org/10.1073/pnas.2406823121
2. **Engelmann, N.,** & **Hannikainen, I. R.** (2024). Does moral valence influence the construal of alternative possibilities?. *Possibility Studies & Society*, 27538699241231026. https://doi.org/10.1177/27538699241231026
3. **Martin, D.,** Díaz-Cobacho, G., & **Hannikainen, I. R.** (2024). The brain death criterion in light of value-based disagreement versus biomedical uncertainty. *The American Journal of Bioethics*, *24*(1), 123-126. https://doi.org/10.1080/15265161.2023.2278566
4. Toomey, J., Lewis, J., **Hannikainen, I. R.,** & **Earp, B. D.** (2024). Advance medical decision-making differs across first-and third-person perspectives. *AJOB Empirical Bioethics*, 1-9. https://doi.org/10.1080/23294515.2024.2336900

2023

5. Flanagan, B., Almeida, G., Struchiner, N., & **Hannikainen, I. R.** (2023). Moral appraisals guide intuitive legal determinations. *Law & Human Behavior,* 47(2), 367-383. https://psycnet.apa.org/doi/10.1037/lhb0000527
6. **Earp, B. D., Hannikainen, I. R.**, Dale, S., & Latham, S. R. (2023). Experimental philosophical bioethics, advance directives, and the true self in dementia. In K. Hens & A. De Block (Ed.). *Advances in Experimental Philosophy of Medicine* (pp. 259–284). London: Bloomsbury Academic. http://dx.doi.org/10.5040/9781350281554.ch-013
7. **Martín, D.**, Rueda, J.**, Earp, B. D.,** & **Hannikainen, I. R**. (2023). Normalization and the treatment/enhancement distinction. *Neuroethics* 16, 13. https://doi.org/10.1007/s12152-023-09519-0.

2022

8. **Hannikainen, I. R.** , Tobia, K., … & Żuradzki, T. (2022). Coordination and expertise foster legal textualism. *Proceedings of the National Academy of Sciences*, 119(44). https://doi.org/10.1073/pnas.2206531119
9. **Demaree-Cotton, J.,** & Sommers, R. (2022). Autonomy and the folk concept of valid consent. *Cognition*, 224, 105065. https://doi.org/10.1016/j.cognition.2022.105065

10. **Bruce, L., Hannikainen, I. R., & Earp, B. D**. (2022). New findings on unconsented intimate exams suggest racial bias and gender parity. *Hastings Center Report*, *52*(2), 7-9. https://doi.org/10.1002/hast.1349
11. **Awad, E.,** Levine, S., Anderson, M., Anderson, S. L., Conitzer, V., Crockett, M. J., ... & Tenenbaum, J. B. (2022). Computational ethics. *Trends in Cognitive Sciences*, *26*(5), 388-405. https://doi.org/10.1016/j.tics.2022.02.009

2021

12. **Earp, B. D.**, McLoughlin, K. L., Monrad, J. T., Clark, M. S., & Crockett, M. J. (2021). How social relationships shape moral wrongness judgments. *Nature Communications*, *12*(1), 5776. https://www.nature.com/articles/s41467-021-26067-4
13. **Everett, J. A.,** Colombatto, C., Awad, E., Boggio, P., Bos, B., Brady, W. J., ... & Crockett, M. J. (2021). Moral dilemmas and trust in leaders during a global health crisis. *Nature Human Behaviour*, *5*(8), 1074-1088. https://www.nature.com/articles/s41562-021-01156-y

Before 2021

14. Kahane, G., **Everett, J. A.**, Earp, B. D., Caviola, L., Faber, N. S., Crockett, M. J., & Savulescu, J. (2018). Beyond sacrificial harm: A two-dimensional model of utilitarian psychology. *Psychological Review*, *125*(2), 131. https://doi.org/10.1037/rev0000093
15. **Awad, E.**, Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59-64. https://www.nature.com/articles/s41586-018-0637-6
16. **Francis, K. B.**, Terbeck, S., Briazu, R. A., Haines, A., Gummerum, M., Ganis, G., & Howard, I. S. (2017). Simulating moral actions: An investigation of personal force in virtual moral dilemmas. *Scientific Reports*, *7*(1), 13954. https://doi.org/10.1038/s41598-017-13909-9

3.6. Recursos humanos, materiales y de equipamiento disponibles para la ejecución del proyecto. ***Human, material and equipment resources available for the execution of the Project*.**

Various resources to which our team has access will be devoted to the project and contribute to its execution: work computers equipped with open-source tools such as jsPsych and Python and access to online platforms like Cognition.run and Qualtrics for study deployment, under active licenses. Additional licenses and equipment are requested and budgeted in Section 5 below. Additionally, our team is composed by a solid international network with a long history of fruitful collaborations and high-impact production in topics that relate closely with the objectives of the MSIMLab project. This ensures that our research objective is feasible and supported by competitive human and material resources.

4. IMPACTO ESPERADO DE LOS RESULTADOS. *EXPECTED IMPACT OF THE RESULTS.*

4.1. Impacto esperado en la generación de conocimiento científico-técnico en el ámbito temático de la propuesta. ***Expected impact on the generation of scientific-technical knowledge in the thematic area of the proposal*.**

This project aims to lead the development of an integrative theory of mental state inference in the moral domain. Drawing on inverse planning models previously applied in basic psychological contexts (e.g., studying inferences about the semantic meaning of a word or the motion of objects) to understand moral cognition (e.g., offensiveness, epistemic injustice, medical decisions) is uncharted territory within cognitive science and will undoubtedly deepen our understanding of how people make moral judgments, both to retrospectively evaluate others and to prospectively decide what to do. As such, the MSIMLab project represents an important contribution to **interdisciplinary research** in the social sciences and humanities—

which we hope will bring about important contributions to the fields of moral psychology, experimental philosophy and cognitive science. The **results** of the project could translate into high-impact publications, participation in national and international events, and spur future multinational collaborations.

4.2. Impacto social y económico de los resultados previstos. *Social and economic impact* of *the expected results*.

This research can be of **significant value to society** by providing a deeper understanding of reasoning processes in high-stakes contexts such as medical diagnosis or judicial verdict. Understanding such reasoning may help to identify cognitive biases and improve decision-making more generally (to enable a fairer and more efficient application of justice in medicine). Additionally, characterizing these inferential processes may be useful in developing more transparent and explainable medical or legal artificial intelligence (**AI)**. Specifically, developing more robust cognitive models of moral valuation can guide ***neurosymbolic approaches*** to AI that implement the idealized cognitive processes that humans use in decision-making contexts (instead of relying on black-box methods, such as deep neural networks).

In this regard, the project's objectives align with **Sustainable Development Goals 16** (*Peace, justice and strong institutions*) and **10** (*Reducing inequalities*) by providing tools to represent mental attributions in validated cognitive models. With regard to **social inclusion and gender**, the project will investigate how these reasoning processes vary along sociocultural and demographic dimensions (WP3), thus contributing to the understanding of the unique challenges that diversity imposes on collective and institutional decision-making.

4.3. Plan de comunicación científica e internacionalización de los resultados (indicar la previsión de publicaciones en acceso abierto). *Plan for scientific communication and internationalization of the results (indicate the forecast of open access publications)*.

The project foresees a holistic **scientific communication and internationalization strategy** aimed at maximizing impact and disseminating the results. Two interim workshops will be organized throughout the course of the project. These will focus on Work Packages 1 (second year) and 2 (third year), with the aim of sharing progress and receiving feedback. Additionally, in the final phase of the project (fourth year), an international conference will be held at the University of Granada, with the participation of recognized experts in related disciplines. The conference will include two or three keynote speakers and a roundtable discussion between cognitive scientists, legal scholars, moral philosophers and non-academic stakeholders (including policy makers and legal practitioners), and will focus on developing the primary theoretical review piece.

The team will also submit ongoing work and give yearly presentations at major international research conferences, foreseeably at the following events: *Society for Philosophy & Psychology, Cognitive Science Society, Society for Experimental Social Psychology, IVR World Congress and ACM Conference on AI, Ethics and Society.* Dissemination of the project will not be limited to specialist conferences but will also seek to present the results in interdisciplinary forums to ensure greater projection and connection with different fields of knowledge. As these findings have significant implications for wider societal issues, the project will also target general scientific journals (e.g. *Science, Nature Human Behaviour*) that have wider implications for civil society.

The project contributes to a vibrant **network of international collaborations**. Highlights include the 2nd International Bioxphi Summit (June 2022, University of Oxford, UK), the 3rd Yale-Oxford-Jagiellonian Workshop on Experimental Philosophical Bioethics (June 2023, Jagiellonian University, Poland), and the forthcoming 4th Workshop on Philosophical Experimental Bioethics (June 2025, University of Basel, Switzerland). These events continue to foster global collaboration, strengthening ties with North America (Yale University) and Asia (National University of Singapore) through recurring meetings.

The MSIMLab project also has a strong **internationalization plan**, including future **collaborations with international R&D groups**. One notable initiative is a joint proposal to the COST (European Cooperation in Science and Technology) program, focusing on the neural substrates and computational basis of moral judgment and mental state inference. This will involve collaborations with key institutions and researchers such as the Research Centre in Applied Ethics at the University of Bucharest (Romania), the Moral Psychology and Experimental Bioethics Lab at the University of Oxford, with the involvement of Prof. Kathryn Francis from the University of Leeds and Dr. Neele Engelmann from the Max Planck Institute for Human Development. These partnerships broaden the interdisciplinary scope of the project, bringing together experts from across Europe and increasing the scientific and societal impact of the research.

The project will commit to publish a minimum of **five open access articles per year**, for an expected total of **20+ open-access publications** over the course of the project. These articles will be published in high-impact journals in fields such as moral psychology, experimental philosophy and cognitive science—thereby contributing to open science.

4.4. Plan de divulgación de los resultados a los colectivos más relevantes para la temática del proyecto y a la sociedad en general. *Plan for dissemination of the results to the most relevant groups for the theme of the project and to society in general.*

The plan for disseminating the results of the project will be designed to reach both the most **relevant groups and society at large**, ensuring that the results have an impact not only on the academic community but also on the general public. The project results may be particularly useful to non-academic professionals, such as policymakers and legal practitioners, who will have a critical involvement in Work Package 3 to help articulate the theoretical implications of our findings for institutional decision-making.

Furthermore, a **dissemination strategy** is planned for the **general public**. The aim is to raise awareness and interest in the issues addressed by the MSIMLab project. To this end, we will produce public-facing blog posts on the project website, as well as on partner sites (FiloLab website of the University of Granada and the Practical Ethics blog of the University of Oxford). We also plan to disseminate press releases and use social media to widen the reach of the project and encourage civic participation and interaction.

To make results more accessible to the international community, non-academic summaries of our main findings will be made available on the project website in several participating languages (through our international network of collaborators).

En los casos en que sea de aplicación: In cases where it is applicable

4.5. Plan de transferencia y valorización de los resultados. *Transfer plan and valorization of results*

Our research has demonstrated the potential for significant real-world impact, particularly in shaping institutional decision-making and public policy. For example, in collaboration with international researchers from Yale and Oxford, a previous research output in the precursor PID2020 project **directly influenced a major policy shift in** the United States (Harrison, 2024): the Department of Health and Human Services' directive QSO-24-10-Hospitals (https://www.cms.gov/files/document/qso-24-10-hospitals.pdf) mandated that hospitals must obtain written informed consent before performing intimate examinations on anesthetized patients. This landmark policy change protects patient rights and underscores the critical role of aligning public policies with citizens' values and concerns.

In this project, we aim to replicate this impact by engaging directly with judicial bodies and political organizations. In the past, I have facilitated workshops with federal judges in São Paulo and Rio de Janeiro, fostering productive dialogues between scholars and legal professionals. These interactions highlighted how empirical research on judgment and decision-making can improve institutional practices. Through similar initiatives, we plan to promote our findings to stakeholders in the judiciary and policymaking sectors, ensuring that our research contributes to evidence-based, ethically informed decision-making processes.

Our transfer plan includes publishing open-access articles to maximize accessibility, presenting results at interdisciplinary conferences, and organizing collaborative workshops to engage decision-makers. Potential stakeholders include judicial councils, medical ethics boards, governmental bodies, and advocacy organizations committed to improving institutional decision-making aligned with public values.

4.6. Resumen del plan de gestión de datos previsto. *Summary´s management plan of the* planned *data*

The MSIMLab project is committed to cumulative knowledge, reproducibility and scientific transparency, adopting best practices in open science (sharing raw anonymized data, stimuli and analysis scripts) in line with the European Union's policy priorities. Second, stimulus materials, anonymized raw data (*.csv*), and analysis scripts (*.r*/*.py*) will be made publicly available on the *Open Science Framework* (OSF) and/or *GitHub* platforms, as we have done continuously over the years (see https://osf.io/m5rgk/). This ensures that other researchers will be able to double-check the reported results and conduct further analyses with which to inform their own work.

In compliance with the open access guidelines for research funded by the Spanish Ministry of Science and Innovation, scholarly outputs—conference presentations and articles—will be published in open access form and made available in preprint form on a dedicated project website. The project website will also host our anonymized datasets, and the validated RoW stimulus set, with a CC-BY license for other researchers to freely re-use and adapt, aiding the field's transition toward more psychologically realistic stimuli.

The data generated and used in the MSIMLab project originate from studies and questionnaires administered to participants through online platforms such as Netquest and Prolific. These datasets undergo cleaning and analysis in the R programming environment, with storage in CSV file format. To comply with FAIR principles, the project implements the following guidelines:

- **Findable:** The data generated in the project are stored on the Open Science Framework (OSF), which assigns a persistent identifier (DOI) to each dataset. This practice ensures that data remain locatable and accessible, regardless of changes to the storage infrastructure.
- **Accessible:** The data are publicly available and hosted on OSF, ensuring that anyone can access them online. Each dataset's associated publication includes a direct link to OSF, facilitating accessibility and usability. A Creative Commons CC BY license permits third-party (re)use of the data, provided proper attribution is given. Accessibility also extends to the required software, as CSV files can be read using general-purpose tools such as Excel, Numbers, Google Sheets, R, and Python.
- **Interoperable:** The data are stored in the widely compatible CSV format, which is easily processed by numerous platforms and software applications. This ensures that the data can be integrated and utilized across a broad range of analysis and processing tools, promoting interoperability between systems.
- **Reusable:** Under the Creative Commons CC-BY 4.0 International Public License, the data generated by the project are fully reusable by other researchers and the public. This license allows complete reuse of the data, enabling reanalysis and replication in future studies, provided the original source is credited. Public access to the data will be enabled once the corresponding article presenting the results is accepted.

Responsibilities and Preservation: The MSIMLab research team is responsible for managing and providing access to the data. All datasets associated with the MSIMLab project will be published and indexed on the Principal Investigator's OSF profile indefinitely: https://osf.io/m5rgk/. The creation of persistent identifiers ensures not only data preservation but also linkage to the published scientific articles.

Software Reuse: To process the data, users may utilize any software compatible with CSV files. This includes open-source tools (e.g., R and Python) as well as commercial software (e.g., Microsoft Excel), maximizing the accessibility and reusability of the generated data.

4.7. Efectos de la inclusión de género en el contenido de la propuesta. *Effects of gender inclusion in the content of the proposal*

The MSIMLab project integrates gender analysis through various approaches. In the methodology**,** we will use **(1) random assignment of masculine and feminine gendered pronouns** in study designs involving individual actions to avoid gender bias. **(2) Prioritization of gender-neutral names** such as "Robin," "Taylor," "Noe," or "Gael" to avoid gender stereotypes (see Hannikainen et al., 2024). **(3) Non-binary gender classification:** Non-binary categories will be adopted in the demographic sections of studies to reflect inclusivity. In participant recruitment, active measures will be taken to ensure **(4) gender-balanced samples** using the "Balanced Sample" option provided by platforms such as Prolific.

In study topics and analyses, there is the possibility that gender differences will arise, as has occurred in a number of past areas of cognitive science research. For example, Costa and colleagues (2019) demonstrated that gender stereotypes concerning warmth and competence play a role in decisions about child custody. Therefore, understanding the **(5) existence and origin of gender differences** might yield important insights into the mechanism by which women are discriminated against or stereotyped in moral decision-making.

During the project, multiple actions will be implemented to foster gender equality in the team and associated activities. (6) Women constitute **50% of the research team**, establish gender parity. (7) **Efforts to increase female participation** will be made in al planned workshops and

collaborations. **(8) Task and role allocation** will be distributed equitably based on individual skills and competencies, avoiding gender biases. **(9) Recruitment committees** will follow gender parity policies.

These nine actions will improve gender disparities within the field, both in team composition and representation in scientific output. The project's focus on inclusivity and equitable practices ensures alignment with broader institutional and societal goals of promoting gender equality in research and in higher education more broadly.

5. JUSTIFICACIÓN DEL PRESUPUESTO SOLICITADO. *JUSTIFICATION OF THE REQUESTED BUDGET*

This research project's objective of unfolding how inference of others' mental states relates to one's moral judgments of their actions will be mainly addressed by means of the collection and analysis of empirical data that can refute or support different metaethical hypotheses. In this sense, each of the three expected WPs will require the conduction of several behavioral web-based experimental studies. With the help of online platforms to collect our study samples (mainly https://www.prolific.com/, but also others such as https://www.netquest.com/, or https://www.bilendi.es/), we expect to conduct approximately 20 experiments for each WP—including pilots and representative surveys. With an estimate of €650 per study, this amounts to a total of **€39,000**, which comprises €30,000 as direct compensation to respondents at a rate of €10 per hour—in compliance with research ethics standards—and €9,000 in fees for the platforms in administrative fees. We also expect to require the acquisition of hardware and software materials for a total of **€14,000**, which include: €2,000 for a dedicated computer, €10,000 for software licenses (hosting and deploying web studies through https://www.heroku.com/ and https://www.cognition.run/), and €2,000 to rent supercomputer hours, primarily tensor processing units, (e.g., through https://colab.research.google.com/) in order to carry out the more computationally taxing modeling and analyses in WP2 (e.g., Monte Carlo Markov Chain sampling in Bayesian forecasting models).

To assist with implementation of web-based experiments, data collection and analysis, and the coordination of cross-cultural replication efforts in WP3, a full-time research Assistantship will be offered. At a total cost of **€39,949**, this will allow us to hire a graduate in Philosophy or related disciplines for a period of 18 months.

Lastly, to foster dissemination and impact of our project, we request **€12,000** to cover the article processing fees in open-access journals (if the University library is unable to cover these expenses through prior agreements). Additionally, travel costs for an estimated total of **€33,600**—at a rate of €8,400 per year, including 6 yearly conference trips for UGR project members at €700 each, plus 2 yearly visits from international scholars to the University of Granada. Thus, these funds would be destined to either travel to conferences and professional meetings or to host external colleagues and invite keynote speakers to the closing conference.

6. CAPACIDAD FORMATIVA. *TRAINING CAPACITY*

6.1. Programa de formación previsto en el contexto del proyecto solicitado. *Training program planned in the context of the requested project*

The PhD candidate will join the PhD Program in Philosophy in the Department of Philosophy I (https://doctorados.ugr.es/filosofia/) of the University of Granada, an institution with a strong track record in philosophical training and research. Their training will be supported by the FiloLab Unit of Excellence (https://wpd.ugr.es/~filolab/wordpress/), a vibrant interdisciplinary

environment that addresses key issues of public debate, combining philosophy with areas such as argumentation, applied ethics and cognitive science.

The PhD candidate will be recruited based on mutual interests, and competence in experimental philosophy and/or moral psychology. They will be actively involved in the research of this project and take lead authorship on one yearly publication. He/she will benefit from all the opportunities of meeting with international experts working in this field, and their research will receive feedback from all team the Advisory Board, and other international experts who have confirmed their willingness to collaborate. The PhD candidate will go on two visiting fellowships of three to four months in world-class moral psychology and experimental philosophy labs, such as the Knobe Lab at Yale University and the Moral Cognition Lab at Harvard University. The research visits will be carried out at either the *Max Planck Center for Minds and Machines*, the *Toulouse School of Economics*, the *Harvard Moral Psychology Lab*, or the *Georgetown Law & Language Lab*.

6.2. Tesis realizadas o en curso en el ámbito del equipo de investigación (últimos 10 años). *Thesis completed or in progress within the scope of the research team (last 10 years).*

Our research team has valuable experience in the supervision of doctoral dissertations that align with the themes of the MSIMLab proposal, reflecting our expertise and ability to guide research in this domain. Below, I provide details of one successfully defended doctoral thesis and three ongoing PhD theses, including the names of the doctoral candidates, their thesis titles, start dates and (expected) completion dates, and funding sources. This information highlights our team's commitment to advancing knowledge in experimental philosophy, moral psychology, and cognitive science, supported by high-quality mentorship and impactful research outputs.

One (1) completed PhD supervision at the Pontifical Catholic University of Rio de Janeiro:

1. Dr. Guilherme Almeida. Institution: Pontifical Catholic University of Rio de Janeiro (enrolled Sep/2017; defended Dec/2020). Title: *Breaking rules: An experimental investigation of the concept of rule.* Co-advised with Prof. N. Struchiner.

Three (3) ongoing PhD students at the Universities of Granada and of the Basque Country:

2. Jorge Suárez (enrolled Sep/2022; expected Sep/2026). University of Granada. *Title: La percepción de los valores ajenos en contextos migratorios y de integración y sus implicaciones ético-políticas*. Keywords: *experimental ethics of migration, multiculturalism.* Funded by a Formación de Personal Investigador (PRE2021-097275) Grant.
    a. **Suárez, J.** & Hannikainen, I. R. (in press). "Limited associations between adoption of host societal values and integration in a sample of twenty thousand migrants". *Frontiers in Sociology*.
    b. Hannikainen, I. R., **Suárez, J.**, Espericueta, L., Menéndez-Ferreras, M. & Rodríguez-Arias, D. (2024). Legal provisions on medical aid in dying encode moral intuition. *Proceedings of the National Academy of Sciences*, 121(42). https://doi.org/10.1073/pnas.2406823121
3. Daniel Martín (enrolled Dec/2023; expected Dec/2027). Institution: University of Granada. Title: *Scope and priority in moral judgement: an experimental approach to the relevance of animals and nature in decision-making*. Keywords: *moral psychology of animal and environmental rights*. Funded by the Fundación Ramón Areces.

a. **Martín, D.**, Rueda, J., Earp, B. D., & Hannikainen, I. R. (2023). Normality and the treatment-enhancement distinction. *Neuroethics*, *16*(2), 13. https://doi.org/10.1007/s12152-023-09519-0

b. **Martín, D.**, Díaz-Cobacho, G., & Hannikainen, I. R. (2024). The brain death criterion in light of value-based disagreement versus biomedical uncertainty. *The American Journal of Bioethics*, *24*(1), 123-126. https://doi.org/10.1080/15265161.2023.2278566

4. Mikel Asteinza (enrolled Jan/2024; expected Jan/2028). University of the Basque Country (co-advised with Prof. J. Umérez). Title: *Condicionantes epistémicos de la divulgación científica y su impacto en la audiencia lega: el caso de la des-extinción.* Keywords: *epistemic risks of popular science*. Funded by a Formación de Personal Universitario (FPU22/03446) Grant.

Desarrollo científico o profesional de los doctores/as egresados/as. *Scientific or professional development of graduate doctors*.

Dr. Guilherme de Almeida transitioned from leading the *Supremo em Números* project at the Getulio Vargas Foundation to a position in Yale University's Cognitive Science Program. He now holds a permanent position as Assistant Professor of Law at Insper Institute of Education and Research in São Paulo, Brazil. His dissertation work resulted in co-authorship of two impactful publications (Struchiner et al., 2020; Almeida et al., 2023).

Dr. Carme Isern-Mas, who held Juan de la Cierva Formación and Margarita Salas awards under my supervision at the University of Granada (2022–2023), secured a position as Assistant Professor of Philosophy at the University of the Balearic Islands. Her postdoctoral research at the University of Granada gave rise to a publication on the folk concept of self-deception (Isern Mas & Hannikainen, 2024), and other work in progress.

These placements underscore the potential for the MSIMLab project to foster high-caliber researchers who contribute significantly to their fields and go on to obtain professional stability.

7. CONDICIONES ESPECÍFICAS PARA LA EJECUCIÓN DE DETERMINADOS PROYECTOS. *SPECIFIC CONDITIONS FOR THE EXECUTION OF CERTAIN PROJECTS*

The studies will be computer-based questionnaires, including self-assessments and other simple judgment tasks. For example, study participants may be asked to view a stimulus—either text, audio or images—and make a judgment about whether the action is moral or immoral, legal or illegal, describing their emotional reaction to stimulus, or judging other unobserved features of a stimulus (who caused an event, what the agent's beliefs or desires were, etc.).

The project involves the participation of human subjects in online and laboratory studies. The following relevant ethical issues have been identified. We must ensure that:

1. participants are informed of the conditions of the research to be conducted,
2. informed consent is granted by participants,
3. participants' data are securely stored.

The study designs, materials and procedures to be performed will be subject to ethical approval by the Research Ethics Committee at the University of Granada (https://investigacion.ugr.es/informacion/presentacion/apoyo/comite-etica). As part of this procedure, the team researchers must describe the study's objectives and protocol, attach an informed consent form (describing how free and informed consent will be sought from all research participants prior to the study) and a non-disclosure and responsible declaration.

Project activities will be conducted in compliance with national and regional legislation on the use of human data for scientific purposes (e.g., in Spain, the Organic Law 3/2018 of Personal Data Protection and Guarantee of Digital Rights), and with the European Parliament's General Data Protection Regulation (GDPR, 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data) which came into effect on the 25th of May 2018, and with any amendments that may be enacted during the project's lifetime.

References

Almagro, M., Hannikainen, I. R., & Villanueva, N. (2023). Who has a free speech problem? Motivated censorship across the ideological divide. In D. Bordonaba Plou (ed.), *Experimental Philosophy of Language: Perspectives, Methods and Prospects*. Springer.

Almeida, G., Struchiner, N., & Hannikainen, I. R. (2023). Rule is a dual character concept. *Cognition*, *230*, 105259.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329-349.

Costa, L. L. F., Esteves, A. B. D., Kreimer, R., Struchiner, N., & Hannikainen, I. (2019). Gender stereotypes underlie child custody decisions. *European Journal of Social Psychology*, *49*(3), 548-559.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084), 998-998.

Goldman, A. I. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Harrison, R. (April, 2024). Yale-led study spurs federal action: HHS requires consent for intimate medical procedures. *Yale ISPS*. https://isps.yale.edu/news/blog/2024/04/yale-led-study-spurs-federal-action-hhs-requires-consent-for-intimate-medical

Hannikainen, I. R., Machery, E., Rose, D., Stich, S., Olivola, C. Y., Sousa, P., ... & Zhu, J. (2019). For whom does determinism undermine moral responsibility? Surveying the conditions for free will across cultures. *Frontiers in Psychology*, *10*, 2428.

Hannikainen, I. R., Tobia, K. P., De Almeida, G. D. F., Donelson, R., Dranseika, V., Kneer, M., ... & Struchiner, N. (2021). Are there cross-cultural legal principles? Modal reasoning uncovers procedural constraints on law. *Cognitive Science*, *45*(8), e13024.

Hannikainen, I. R., Tobia, K. P., de Almeida, G. D. F., Struchiner, N., Kneer, M., Bystranowski, P., ... & Żuradzki, T. (2022). Coordination and expertise foster legal textualism. *Proceedings of the National Academy of Sciences*, *119*(44), e2206531119.

Isern-Mas, C., & Hannikainen, I. R. (2024). Self-deception: A case study in folk conceptual structure. *Review of Philosophy and Psychology*, 1-28.

Margoni, F., & Surian, L. (2016). Mental state understanding and moral judgment in children with autistic spectrum disorder. *Frontiers in Psychology*, *7*, 1478.

Pole, A., West, M., & Harrison, J. (2018). *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC.

Struchiner, N., Hannikainen, I. R., & de Almeida, G. D. F. (2020). An experimental guide to vehicles in the park. *Judgment and Decision Making*, *15*(3), 312-329.

Toomey, J., Lewis, J., Hannikainen, I. R., & Earp, B. D. (2024). Advance medical decision-making differs across first-and third-person perspectives. *AJOB Empirical Bioethics*, 1-9.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, *104*(20), 8235-8240.