Philosophical Intuition and Non-Classical Conceptual Structure

Dr. Ivar Rodríguez Hannikainen ivar@ugr.es
Universidad de Granada www.moralsciencelab.com

Within analytic philosophy, there is a long tradition of appealing to intuitions about thought experiments to probe the structure of core philosophical concepts, such as morality, knowledge or identity. The role of such thought experiments is to elucidate these concepts by testing proposed criteria of necessity and sufficiency against our intuitive judgments. Consider, for example, the Frankfurt case[1]—which asks us to consider whether an agent who lacks alternative possibilities may nevertheless have acted out of free will. The underlying argument is that, if we intuit that the agent *is* indeed free, despite not having alternative possibilities, we will have shown that alternative possibilities cannot be a necessary condition for free will. In this way, this philosophical method, known as the *method of cases*, assumes that intuitions reflect the application of *classically* structured concepts—i.e., concepts with necessary and sufficient conditions.

Meanwhile, a growing empirical literature in cognitive science has shown that many natural and social concepts do not exhibit a classical structure. For example, when reasoning about whether something is a BIRD or a GAME, we do not primarily consider the necessary and sufficient conditions for category membership. Instead, we often compare the target entity to a *prototypical* or *exemplary* BIRD along multiple similarity dimensions (such as whether it flies, nests in trees, or sings). Collectively, these studies have given rise to non-classical theories of concepts[2]. Might intuitions about philosophical concepts, such as morality, free will or identity, also exhibit a non-classical structure?

The *Philosophical Intuition and Non-Classical Conceptual Structure* project asks whether **intuitions about core philosophical concepts** (e.g., knowledge, free will, and identity) **reflect prototype-based reasoning** rather than the application of classical definitions.

This prediction is inspired by three insights: (1) Many philosophical questions have remained hotly contested for centuries. (2) Appeals to commonsense intuition about these questions have rarely illuminated these debates, and instead revealed parallel tensions in laypeople's intuitions[3,4]. (3) These tensions primarily reflect *intra*personal conflict rather than *inter*personal disagreement[5-7] (see also Fig. 1).

By applying three complementary methods from cognitive science (see **Methodology**), the project will (SO1) empirically test whether laypeople's philosophical intuitions on core issues exhibit a non-classical or a classical structure, (SO2) assess how conceptual structure shifts with philosophical expertise, and (SO3) articulate the implications for philosophical methodology, e.g., regarding the epistemic role of intuitions in conceptual analysis.

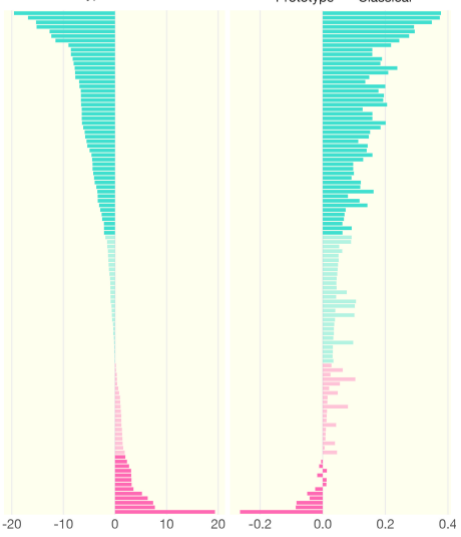

Fig. 1. Most participants exhibit a prototype-based (~70%) and not a classical (~30%) concept of SELF-DECEPTION (adapted from ref. [7]).

Specific Objectives

SO1. Do ordinary intuitions about philosophical concepts (e.g., knowledge, intentionality or identity) reflect prototype-based reasoning?

SO2. What is the effect of philosophical expertise on conceptual structure? Might expertise ingrain a *classical* conceptual structure?

SO3. What are the implications of non-classical conceptual structures for philosophical methodology (e.g., the epistemic status of intuitions)?

Methodology

To meet Specific Objectives 1 and 2, I will conduct a series of empirical studies, triangulating three complementary methods to establish whether core philosophical concepts exhibit a prototypical or classical structure.

Method 1. Typicality ratings.

According to the classical view of concepts, a target should be categorized as an instance of the concept (e.g., a BACHELOR) *iff* it fulfills the individually necessary and jointly sufficient conditions (i.e., being unmarried, adult and male). However, research by Eleanor Rosch[8] and others[2] shows that people judge some instances (e.g., a 28-year old man on Tinder) as more typical than others (e.g., an 80-year old celibate monk), and classify typical instances as category members more often than atypical instances (see also Fig. 2). In Method 1, we will adopt this technique to evaluate whether laypeople treat typical instances of a philosophical concept (e.g., an evil act[9], with alternative possibilities, or with sourcehood[10]) as instances of the concept (i.e., FREE WILL) more so than atypical instances (e.g., an inconsequential act, without alternative possibilities, or without sourcehood).

Method 2. Conflict and response competition.

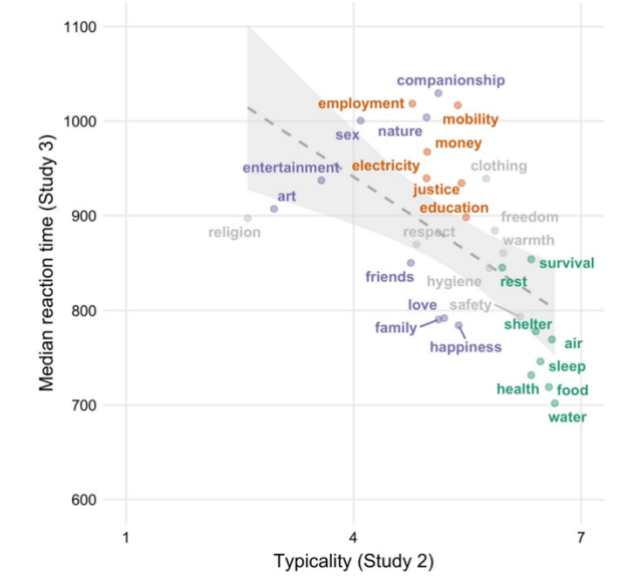

Fig. 2. Typicality ratings and reaction times provide convergent evidence of prototype-based reasoning about concepts, such as BASIC NEEDS (reprinted from ref. [11]).

Past work has also found *faster* classification of prototypical items (e.g., *a robin* as a BIRD) than atypical items (e.g., a *penguin* as a BIRD). In recent studies, we have shown that this pattern extends to philosophically relevant concepts, such as *basic needs*[11] and *rule violation*. For example, prototypical violations of a rule are decided faster than atypical violations involving transgression of the rule's letter but not its spirit, or *vice versa*[12].

Accordingly, we will carry out reaction time studies to ascertain whether people are hesitant in atypical cases of low similarity to prototypical *philosophical* concepts, investigating whether reaction times yield convergent evidence of prototype concept use (as in Fig. 2).

Method 3. Criterion-based vs. similarity-based classification.

Isern-Mas and I[7] recently developed a *modeling* approach to empirically distinguish prototype concept from classical concept use. The proposed approach is to compare the model fit (using AIC and r^2 as in Fig. 1) of *by-participant* regression models based on criteria (see Eq. 1) versus a weighted sum of similarity dimensions (see Eq. 2).

We first fit every theoretically plausible model specifying necessary and sufficient conditions for something to instantiate the concept, *C*. Thus, we ask: If indeed this individual has a classical concept of *C*, which concept is it statistically most likely to be?

[Equation 1]

Criterion-based model

$$prob.\,(\mathrm{C}=1)=\frac{e^{a\times\mathrm{Criterion}+c}}{1+e^{a\times\mathrm{Criterion}+c}}$$

We then repeat this procedure with similarity-based models. We fit every plausible similarity-based model and ask: If this participant has a prototype concept of *C*, which features of the prototype do they attend to when classifying targets as instances of *C*?

[Equation 2]

Similarity-based model

$$prob.\,(\mathrm{C}=1)=\frac{e^{\sum_{i=2}^{m}a_i\times\mathrm{Dimension}_i+c}}{1+e^{\sum_{i=2}^{m}a_i\times\mathrm{Dimension}_i+c}}$$

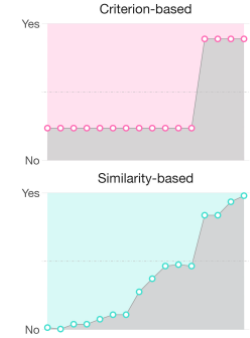

Fig. 3. Example *criterion-based* and *similarity-based* models of SELF-DECEPTION (adapted from ref. [7]).

The final step in Method 3 is to compare—for each participant—the best-fitting criterion-based model to the best-fitting similarity-based model, using *non-nested model selection*. In short, this method allows us to distinguish participants who appear to rely on necessity and sufficiency criteria from those who rely on graded similarity to a prototype[5,7] (see Fig. 3).

Participant Sampling. In Work Package 1, nationally representative samples will be recruited on the crowdsourcing platforms Prolific (for English-speaking countries), Netquest (for Spanish-speaking countries) and Credamo (for Chinese-speaking countries). In Work Package 2, academic philosophers and philosophy students will be recruited using *snowball sampling* through our collaboration networks.

Stimuli and Design. The studies in this project will involve variants of classic thought experiments in philosophy (e.g., Phineas Gage, Ship of Theseus, Twin Earth, or Frankfurt cases). In within-subjects multifactorial designs, our stimuli will orthogonally manipulate candidate features (e.g., sourcehood, alternative possibilities, moral valence) of a concept (i.e., FREE WILL)—generating cases that range from highly prototypical to highly atypical. Translations will be provided by myself (into Spanish) and collaborators at Zhejiang University (into Simplified Chinese). The studies will be forced-choice classification tasks implemented in jsPsych, a Javascript-based framework that enables accurate online measurement of reaction times (besides ratings and responses).

Statistical Analysis. To calculate mean similarity (Method 1) and median reaction times (Method 2) and fit classification models (Method 3), data analysis will be performed in the *R* programming language using the *rstan* and *brms* packages.

Work Plan

I have devised a three-part work plan over the proposed eighteen-month duration of the project to achieve Specific Objectives 1-3 (see Fig. 4).

Work Package 1: The Conceptual Structure of Philosophical Intuition involves a series of web-based experiments with nationally representative samples. WP1 will apply Methods 1-3 to understand whether ordinary philosophical concepts exhibit a prototype structure. Task 1.1a will involve stimuli development and pre-testing of study materials. Task 1.2a will involve 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 (e.g., the *Society for Philosophy and Psychology*, *American Philosophical Association* and *Sociedad Española de Filosofía Analítica*). Task 1.1b will involve finalizing the stimuli and translating to additional languages, and Task 1.2b will involve running the pre-registered experiments with nationally representative samples. In Task 1.3b the manuscript will be updated to include the results obtained in Task 1.2b. Deliverable 1.4 will involve the submission (1.4) and revision (1.4*) of a crosscultural study on the conceptual structure of folk philosophical intuition (e.g., for *Mind & Language*).

Work Package 2: How Philosophical Expertise Shapes Conceptual Structure involves a quasi-experiment on the effect of philosophical expertise. This work package will reuse the study materials developed in WP1 and apply Methods 1-3 to compare philosophers' and laypeople's use of central philosophical concepts. Tasks 2.1a will involve the experiment programming and development of the *Philosophy Academy* website, and creation of the collaborative network. Tasks 2.2a and 2.2b will entail data collection with trained philosophers (2.2a) and a matched group of non-philosophers (2.2b) groups. Tasks 2.3a will involve the preliminary analysis of responses in the group of philosophers for submission to an academic conference, and Task 2.3b will involve the definitive comparative analysis and write-up of the empirical results. Deliverable 2.4 will involve the submission of an empirical report (e.g., to *Cognition* or *Noûs*) examining the effects of philosophical expertise on conceptual structure.

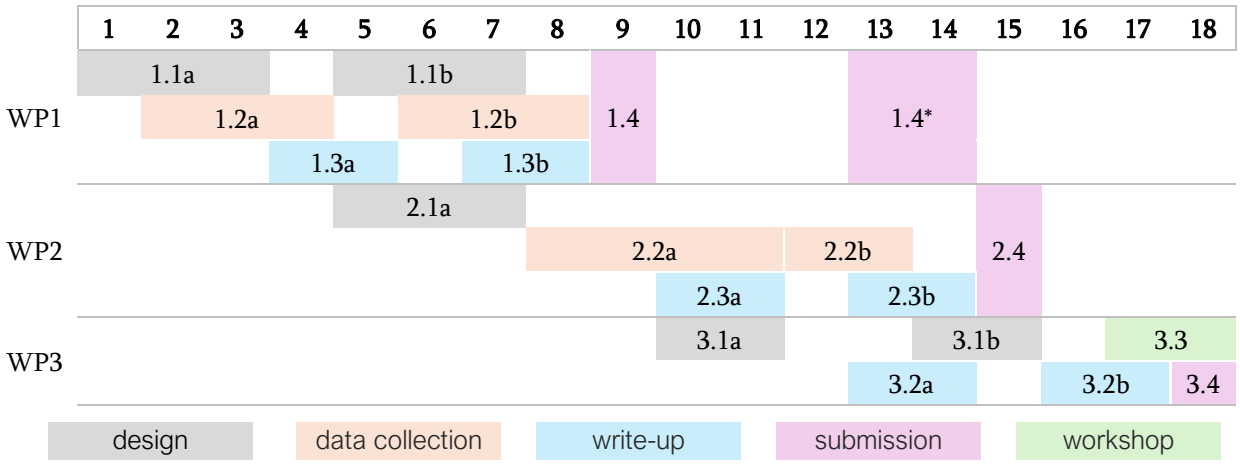

Figure 4. Project Timeline in Gantt Chart Format.

Work Package 3: The Metaphilosophy of Prototype Concepts will invite a core group of collaborators to convene at a closing workshop and co-author an argumentative review paper. In Task 3.1a, we will initiate the preliminaries of the workshop including contacting invited speakers and distributing a call for abstracts, inviting submissions from academics working in related fields (such as cognitive science and linguistics). Task 3.1b will be to evaluate abstracts, finalize the reservations and publicize the event online. Deliverable 3.4 will involve holding the workshop on *Philosophical Intuition and Non-Classical Conceptual Structure* at the University of Granada. The workshop will serve as a platform to disseminate and discuss the results of the project with collaborators and colleagues

from neighboring disciplines. Task 3.2a will involve authoring a first draft of the theoretical paper and distributing it among collaborators for feedback. Task 3.2b will involve integrating coauthors' feedback and revising the theoretical manuscript. Deliverable 3.4 will involve submission of the manuscript to a theoretical outlet (e.g., *Trends in Cognitive Sciences* or *Behavioral and Brain Sciences*).

Contributions and Significance

Can conceptual analyses be fruitfully informed by appeals to commonsense intuition? By building a timely bridge between philosophy and the cognitive science of concepts, this project is poised to make two valuable contributions to this central question in metaphilosophy. First, understanding whether intuitions about central philosophical concepts exhibit classical or non-classical structures would help to adjudicate in the debate about whether conceptual analysis aims to identify necessary and sufficient conditions or uncover similarity dimensions[13,14]. Second, if philosophical training and expertise are found to fundamentally influence conceptual structure, this realization would support a *revisionist* view of conceptual analysis—one in which the primary role of philosophy is not to accommodate intuitions in a parsimonious analysis, but to reorganize and refine them through systematic reflection.

As part of the project's long-term legacy, I will develop the *Philosophy Academy* website—an open-access resource to support university-level philosophy instruction. The platform will feature interactive modules to aid in teaching complex thought experiments while recording visitors' intuitive judgments. In this way, the platform will simultaneously enrich philosophical pedagogy and contribute to experimental philosophy research, offering a rich dataset on students' intuitions about classic and contemporary thought experiments.

References

1. Frankfurt H. Alternate possibilities and moral responsibility. *J. Philos*. **66**, 829–839 (1969).
2. Zaki SR, Nosofsky RM, Stanton RD, Cohen AL. Prototype and exemplar accounts of category learning and attentional allocation: a reassessment. *J. Exp. Psychol. Learn. Mem. Cogn.* **29**, 1160–1173 (2003).
3. Cushman F, Greene JD. Finding faults: How moral dilemmas illuminate cognitive structure. *Soc. Neurosci.* **7**, 269–279 (2012).
4. Knobe J. Conflicting intuitions. *Ergo* (in press).
5. Almeida G, Flanagan B, Hannikainen IR. Trait empathy predicts a preference for the spirit of the law. *J. Res. Pers.* (2025).
6. Dranseika V. Two ships of Theseus. *Synthese* **203**, 1–14 (2024).
7. Isern-Mas C, Hannikainen IR. Self-deception: A case study in folk conceptual structure. *Rev. Philos. Psychol.* 1–28 (2024).
8. Rosch, E. Cognitive representations of semantic categories. *J. Exp. Psychol. Gen.* **104**, 192–233 (1975).
9. Nichols S, Knobe J. Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs* **41**, 663–685 (2007).
10. Hannikainen IR et al. For whom does determinism undermine moral responsibility? Surveying the conditions for free will across cultures. *Front. Psychol.* **10**, 2428 (2019).
11. Pölzler T, Hannikainen IR. The typicality effect in basic needs. *Synthese* **200**, 1–26 (2022).
12. Engelmann N, Hannikainen IR, González-García C, Ruz M. Understanding rule enforcement using drift diffusion models. In *Proc. Annu. Meet. Cogn. Sci. Soc.* **46** (2024).
13. Fodor JA. *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press, Oxford (1998).
14. Stich, S. What is a theory of mental representation? *Mind* **101**, 243–261 (1992).