Contents lists available at ScienceDirect

New Ideas in Psychology

journal homepage: www.elsevier.com/locate/newideapsych

A theory of moral categorization: The conceptual performance of moral cognition

Ariel James

Department of International Relations, Faculty of Humanities and Social Sciences, Universidad Pontificia de Comillas ICAI ICADE CIHS, Calle Universidad Comillas, 3-5, 28049, Madrid, Spain

ABSTRACT

According to some traditional theories of categorization, human beings conceptually represent reality through judgments of similarity between the relevant features of some case, and the well-defined features of the concept that stands for the case. However, abstract concepts used with a moral sense do not have phenomenological resemblance with the cases they denote. In addition, concepts typically deployed in moral judgments are subject to semantic variation. Instead, this article proposes that moral categorization is based on an operational test that verifies whether certain impersonal (generic) and unconditional commands are observed or not in each specific case. The process of moral categorization is described using a cognitive procedure called "categorization by testing the command", consisting of five steps, namely: Descriptive stage, non-normative evaluation, normative question, conditional rule, and categorial construal. Finally, the phenomena of semantic polarity and semantic variation applied to the field of moral cognition are analyzed, to show how the proposed model can address the effects of these interpretive mechanisms.

1. Introduction

Moral cognition can be understood as the set of the bio-neuro-psychological processes that are involved in the generation of moral judgments, decisions, attitudes, and actions.

One of the first questions that comes to mind when defining moral cognition is: What is the basic cognitive structure for the mechanisms of moral evaluation, decision-making, and behavior? It is an inquiry into what differentiates the psychological process of moral evaluation from, say, other forms of cognitive processing such as pattern recognition, arithmetic thinking, or verbal comprehension.

In a very schematic way: What makes moral cognition a well-defined mental state and function within the overall organization of our mind/brain system?

Some interesting answers have been offered from the fields of moral psychology and moral philosophy to solve this fundamental question (See Johnson, 2014; Greene, 2015; Bartels, Bauman, Cushman, Pizarro, & McGraw, 2016; Liao, 2016; Cosmides, Guzmán, & Tooby, 2018).

One of these is the idea that moral evaluation obeys a series of heuristic emotional processes - that are somehow given in the psycho-affective constitution of the species - without implying a connection with more analytical and deliberative aspects of cognition (Haidt & Joseph, 2007). A derivation of this hypothesis holds that the “moral mind” is divided into two well differentiated and autonomous parts, namely, an automatic and emotional system, mainly located in the ventromedial prefrontal cortex (VMPC), as opposed to a more impersonal or rational system, dedicated to producing abstract conscious applications of explicit principles, located in the dorsolateral prefrontal cortex (Greene, Nystrom, Engell, Darley, & Cohen, 2004). Another theory proposes that moral evaluation is the product of an innate digital device for algorithmic/syntactic operations, defined as a "universal moral grammar" (UMG), following Chomsky’s linguistic generative parameters (Hauser, 2006; Mikhail, 2011).

In these approaches it is presupposed that moral categories are axiom-like entities whose existence is self-evident, and therefore, without need of further interpretation and problematization. They usually develop complex and detailed descriptive constructions, but the elemental constituents – the conceptual units of moral evaluation - upon which these buildings are sustained remain unexplored.

The theoretical and empirical problem that those theories do not address, and which leaves our initial question unanswered, consists of understanding how and why the main vehicles of expression of moral cognition, which are the specific value judgments denominated as “moral”, can arise, and develop from a state in which there is no absolute criterion that can set the priority of some values over others. How is it that some values can be selected from among all the others to fulfill the functions of normativity and obligatoriness within a moral point of view? How does the sense of moral oughtness come to be *conceptually* represented?

The concept of “values” refers to abstract behavioral orientations that reflect desirable goals, also known as “terminal values”, and the means to achieve those goals, or “instrumental values” (Rokeach, 2008).

E-mail address: ajames@comillas.edu.

https://doi.org/10.1016/j.newideapsych.2023.101047
Received 14 September 2022; Received in revised form 10 July 2023; Accepted 10 August 2023
Available online 18 August 2023

Values seem to be organized into what some researchers call a “hierarchy”, which works from an evaluation of the relative importance of each value in relation to all the others in a certain context (Rokeach, 2008; Schwartz, 1992; Tetlock, 1986). It seems that outside of moral criteria, that is, regardless of normative ideals such as rightness, equity, or the assumption of moral responsibility, there is no rational way to order the preference of subjective values according to parameters that would not be relative, contingent, and arbitrary – although it could be argued that the hierarchy of values in general cannot be reduced to the domain of moral cognition. To some extent that is not superfluous, if there is no moral (objective, transpersonal) parameter for evaluation, neither people nor human groups need to be coherent in their value orientations – the problem enunciated by Plato in the *Protagoras*.

There are as many value standards as there are different persons and human groups with their own practical necessities, appetites, and interests (Geertz, 1973; Kohlberg & Hersh, 1977). Quantitative empirical research shows that even though it is possible to propose a typology of “universal” motivational values such as benevolence, security, power, and hedonism, among others (Schwartz, 1992), it is the position in the social structure that finally determines the orientation of values of individuals (Kohn & Schooler, 1983). Moreover, the diverse concepts of the good and the countless possible human ends are usually opposed to each other and cannot be subject to a common acceptable measure (Rawls, 1999, p. 150).

It seems that, in the absence of the regulatory role of moral beliefs and normative ideals, there is no rational way to arrange the multiplicity of human values, interests and ends within a well-defined and coherent ranking of priorities. Accordingly, one of the first questions that a theory of moral cognition may try to answer is: How is it possible for optional, non-obligatory and contingent values to become or to be used as values that are obligatory to the extent that they are assumed as moral duties?

In short, it is about knowing how the values that describe an ethically significant event are connected to or are transformed into the moral categories that classify it. Let us call this the *normativity problem*, in other words, the problem of knowing how we build or perhaps perceive and select certain normative, mandatory, and impersonal ethical/moral categories from a non-normative space of subjective values, goals, and ends. This is an acute problem for any theory of moral knowledge, both in psychological, philosophical, juridical, and anthropological terms.

I will avoid the resource to replace the non-normative sense of certain values by terms such as "preference" or "choice" since preferences and choices can also apply to processes of normative evaluation. When two moral duties are in conflict, there is a mechanism of preferential judgment that allows us to choose between them, and this type of "preference" is moral. Therefore, I assume that every evaluative concept has a normative nuance. As I shall distinguish between the “non-moral” and the “moral” connotations of concepts, for the first case I will speak of *non-normative values* (i.e., goals and desires that are not considered to be collective norms), and for the second case I will use the category of *normative values* in general.

Normative values tend to be described as ethical norms – ranging from mere social conventions such as rules of etiquette up to the pragmatic rules of communication that imply a cooperative effort on the part of the speakers such as "Do not say what you believe to be false" (Grice, 1991). In turn, ethical rules can be conditional - for example, those rules covering contractual relationships - or unconditional. In this article I will concentrate mainly on the second option.

Certain ethical norms can be interpreted as if they expressed unconditional mandates holding for each person and in every possible situation, regardless of external constraints and particular interests. When ethical norms are taken in this unconditional sense, we usually refer to them as "moral duties" - e.g., the individual right to self-defense, the principle of non-discrimination, the rule prohibiting objective harm and wrongs, etc. I use the notion of unconditional command to describe all those ethical mandates conveying this sense of duty. Later we will see how unconditional rules are connected to certain conditional clauses, which allows them to be context sensitive.

2. Research question

The research question addressed in this article is the following: What is the cognitive mechanism that allows the production of moral categorization, understood as the process by which people can classify different types of actions, omissions, and their respective repercussions, based on their degree of conformity to impersonal and unconditional obligations and ideals?

In this context, “impersonal” means that the norm has a generic, non-specific, or general reference: it applies to any person, everybody. If you will, it is directed to a 4th grammatical person, comprising the 1st + 2nd + 3rd persons. This fourth person (generic subject/object) is the reference point for moral evaluation and judgment, to the extent that the *most general* unconditional obligations are *not* intended to be tailored to concrete individuals, families, or sub-groups, but rather apply to the entire collective.

In turn, “unconditional” means that one is obliged to respect the norm/duty regardless of how other people act. One is primarily bound to the norm, and then, through it, to individuals.

3. Categorization by testing the command: the basic structure of moral categorization

For the sake of clarity, I will denominate this model of moral categorization as *categorization by testing the command* (CTC). It is a simplified theoretical model with just a few constituents, not a real description of any concrete event. In this section I offer a general idea of the basic constituents that must be considered to operate this model. In the next section I will produce a five-step procedure explaining how it works in a hypothetical case.

The first three constituents are: a real or imaginary *case* (i.e., situation, conflict) that is mentally coded using iconic and symbolic tools, followed by a *conditional rule* that checks if the case respects certain commands (prescribing some unconditional obligations), from which it is possible to decide whether it can be classified with a particular *concept*. To have a picture of the binding between the conceptual notion (type) and its instances (tokens), let us arrange the entire mechanism using just three signs, namely: (a), (b), and (c), in this way:

(a) is the lexical concept, which functions as the abstract attribute that may (or may not) represent the case,

(b) is a mental iconic-like representation of some specific case (person/situation/action), based upon a particular configuration of different kinds of salient semantic features such as agent, patient, action, omission/commission, intentions, causal proximity, consequences, inter alia, whereas

(c) is the rule, or a check mechanism determining whether case (b) can be represented by concept (a), which can only be established *if* the case conforms to or respects an implicit or explicit command. Therefore, the fourth constituent that I shall use is the notion of "command" (throughout the article I offer several definitions of it). To be used in the context of moral evaluation, the command must convey the order to follow an impersonal and unconditional obligation in any situation where a conflict between interest and unconditional duty, or between conditional and non-conditional duties, arises.

The transition between concept (a), and case (b), depends on verifying the accomplishment of a prescriptive command in/by the checking mechanism (c). In other words, only to the extent that the application of the command can be verified by the checking rule (c), is it possible to assume that concept (a) is the appropriate label for representing case (b).

The command is nothing more than a prescription that demands compliance with some normative ideal, that then can be represented by a lexical concept. However, it is difficult to establish an absolute line of demarcation between the notions of “concept” and “command”, insofar as certain concepts such as *fairness* could imply considerations of normativity and obligatoriness. On the other hand, commands necessarily involve conceptual knowledge, and the conceptual knowledge that is embedded in certain explicit commands in turn may range from very abstract rules (e.g., "treat similar cases similarly") to more concrete rules (e.g., "in this case, apply an egalitarian distribution instead of a proportional one").

My proposal is that the activity of moral categorization is the process by which it is verified how close the case is to the command, not to the concept. The concept is secondary, its activation depends entirely on the prior recognition of the applicability of the command. The analytic model suggested here is simple and at the same time explanatory: instead of collecting the set of all the relevant features of the case, to check whether they exhibit some sort of similarity with the set of all the relevant features of the concept, it is enough to check whether the command conveying the ideal of conduct holds or at least is not violated in the case.

The comparison is not between the features of the action and the features of the concept, but rather between the mental representation of action and its degree of compliance with the command. The command is an abstraction of all the possible features of some impersonal and unconditional duty - of course, it is a heuristic device. In turn, the lexical concept itself is just a *formal* representation, since it can convey opposite meanings - that is, both moral, non-moral, even neutral senses. As such, the specific “moral” meaning of concepts used in moral cognition is supported by the set of normative ideals regarding impersonal-unconditional obligations (understood not only as norms, but also as ideals).

Basically, the checking mechanism consists of knowing if these normative ideals are satisfied or not in a case, which implies a special sensitivity for incidents that violate the pragmatic rules (Cheng & Holyoak, 1985; Holyoak & Cheng, 1995; Kruglanski & Gigerenzer, 2011). In some instances, the command *directly* orders some type of action to be carried out or to be avoided. However, in other instances, even when there is no overt obligation to perform certain action or to avoid it, both the acts of performance and avoidance must respect the demands of the set of commands and imperatives. If we assume a pragmatic perspective, commands are both internal and external: they express an internal frame of mind (a parameter for judgment), while they communicate through different culturally codified linguistic modalities such as direct mandates, and the more polite formulas of advice, requests, invitations, and supplications.

Therefore, commands used within moral evaluations are not just explicit orders but constitute the framework of constraints that regulate human behavior in general - applying both to mandated and to optional choices and actions, both to sacred and profane spaces (Cf. Atran, 2021). If the command happens to be respected in some specific situation, the positive “moral” concept is the appropriate label for representing the case. On the contrary, if the conditional rule verifies that the command that orders an unconditional duty is not respected/observed in practice, then the positive concept may be replaced by its opposite pole, loaded with a negative valence – e.g., as in fairness versus unfairness.

Between these two opposite semantic poles it is possible to assume that there may be a variety of intermediate evaluative options, since moral categorization is not a discrete process of reasoning, but a continuous and graded one (James, 2017), with many different nuances (e.g., as in the case of those adjectives and adverbs that become comparative and superlative). Cillian McHugh and its colleagues quite rightly point out that some cases may be "morally irrelevant" (McHugh, McGann, Igou, & Kinsella, 2022), which I suspect gives us two possibilities: either the moral judgment is suspended in those cases, or rather there are specific judgments for neutral cases.

The *categorization by testing the command* model (CT) proposed here suggests that the basic structure of moral categorization does not consist in knowing whether the category is somewhat “similar” to the characters and events that it represents. Some traditional theories of categorization hold that a case can be classified as an instance of a candidate category to the extent that it has the necessary and sufficient “properties” that define the category in question – a position that is usually attributed to Aristotle’s *Metaphysics* (Cf. Taylor, 1995, pp. 22–24), although the *Organon* gives us a less absolute picture.

Other models like “exemplar theory” defend the idea that the concept is a sort of collection that groups all the instances of similar objects or actions within a specific memory space (Kruschke, 1992), but this theory was clearly not intended for those cases in which one knows in advance that torturing a living being is unconditionally wrong, without ever having been exposed to such experience. Furthermore, other approaches presuppose the existence of a prototypical instance whose relevant and pertinent features are compared with the features of the case, to decide whether the case fits the prototype (Rosch, 1975).

All these theories depend on judgments of similarity between the features of the case and the features of some conceptual representation – for a schematic presentation of their differences in relation to moral reasoning see Harman, Mason, and Sinnott-Armstrong (2010).

Although these approaches may be useful to understand certain aspects of moral cognition, they are incomplete to truly explain it: moral evaluation ultimately depends on the possibility of verifying whether specific commands are respected or not in each case. Categorizing an event as moral or immoral is the result of a cognitive process that verifies degrees of compliance with the norm - everything else is tangential and may be irrelevant to the extent that there is no phenomenological or sensory-motor similarity between an instance and its correspondent conceptual representation. According to the approach adopted here, abstract categories are neither reducible nor derivable from perceptual or functional features (though see Barsalou & Wiemer-Hastings, 2005).

A successful act of basic moral categorization - whether classified with a positive categorization judgment such as in "the *correct* thing to do", or with a negative one such as in "the wrong thing to do" - depends on a cognitive operational test that functions through a conditional rule, checking whether a particular case comply or does not comply with implicit or explicit mandates, that are based on certain collective models of behavior (i.e., impersonal-unconditional duties understood as normative ideals), that then can be represented by lexical concepts such as justice, benevolence, mutual aid, and mutual respect, among others.

In turn, understanding the context is crucial to know when an axiological notion like "benevolence" expresses a sense of obligation that is conditional, or rather unconditional - i.e., moral theory is not separated from pragmatic knowledge. Basically, different types of "obligations" are linked to different types of contexts. The delicate balance between conditional and unconditional duties can be tipped on either side, and whenever that crucial ethical distinction is blurred the consequences become problematic, to put it mildly. E.g., in some contexts, notions such as "purity", or personal sacrifice, are demanded as unconditional obligations, which cannot fail to be unfair to all those who suffer the consequences (the classic problem reflected in Iphigenia).

Accordingly, the basic structure of moral categorization can be construed as the process by which certain categories are applied to particular cases to the extent that two fundamental prerequisites are satisfied: first, the case must conforms to an ideal of impersonal-unconditional duty, and second, we know that it conforms to this ideal *because* it complies with certain specific commands, that apply to the open set of “persons in general” (i.e., to a generic person), sometimes trying to get close to universal quantification (∀).

Although it is true that the process of moral evaluation does not start from scratch each time, but refers to a background knowledge - e.g., McHugh et al. (2022) argue very persuasively in their theory of “moral judgment as categorization” (MJAC) that "relative stability in moral categorizations emerges as a result of continued and consistent type-token interpretation such that particular categorizations become habitualized (and hence intuitive)" (Idem, 139) – notwithstanding, the factor that explains the phenomenon of moral conceptualization-generalization is not the habit *per se*, but rather the cognitive mechanism that is progressively internalized, until it becomes a routine for some stereotypical situations. In short, what governs the moral categorization process is not a habit based on statistical frequency, extensibility, and repeatability, but rather a check mechanism for testing conformity to commands, based on a conditional rule.

4. The checking mechanism step by step

Next, I will use a hypothetical case to show how the checking mechanism operates. Imagine the following scenario,

- A pedestrian crosses the crosswalk without realizing that the traffic lights have changed, and the "don't walk" signal has been activated. Immediately after the signal appears, a driver decides to speed up and runs over the pedestrian who was about to get on the sidewalk. Gathered several witnesses to the event, the following statements emerge:

(I). *Descriptive stage.* Statement by which the facts are described: "Individual X, on such a day, at such an hour, driving on road Y, performed the following act … ".

(II). *Non-normative evaluation.* "The driver declares that he/she did that action compelled by a special frame of mind (stress, emotional instability, etc.), without paying much attention to the traffic situation, etc."

(III). *Normative question* – "normative" in the sense of an impersonal and unconditional duty - that triggers the process of *moral* categorization:

"*Should* the action described in point (I) be carried out, regardless of the possible "motives" (e.g., stress level, emotional states, lack of attention, and so on) alleged in point (II)?"

A brief hiatus regarding point (III): Unlike traffic rules, which are civic regulations whose binding is conditional to the respect of superior unconditional norms - such as the prohibition of harms and wrongs - the command activated by the "should" is at the same time binding and unconditional. It does not derive from or depend on the set of traffic rules: it is valid and justified above and despite them. What this model (CTC) suggests is that the set of unconditional obligations is superior to the set of conditional mandates, even though both levels are necessary if it is about having a minimally functional social structure.

Question (III) refers directly to a command that prescribes what should and shouldn't be done. The command (i.e., "You should not do it") applies to all situations of the type represented in the description. Note that the command is an indispensable constituent within the moral categorization process: without its mediation there is no practical or logical way to find out how many types should be assigned to each token - the list can be as arbitrary as unbounded (the problem that moral relativism tries to exploit).

The command in the second person has been usually described as an "imperative" construction - especially after Kant - however, natural language shows great versatility in this regard: an unconditional mandate can also be expressed using conditional formulas, such as the deontic conditional "Even if you do it, it is not all right" (See Clancy, Akatsuka, & Strauss, 1997, 24). Moreover, certain sentences expressing an explicit deontic prohibition also contain a non-deontic conditional indicating the possible consequences of carrying out the prohibited action (Idem, 25) - which allows us to assume that unconditional duties are something more than just imperative formulas. In any case, the sense of obligation is not reduced to the imperative mood.

(IV). *Conditional rule*: *If* the answer to question (III) is negative – i.e., "the action should not be carried out *because* it violates the command" - *then* the categorization judgment will be negative. In this way it is feasible to assume that the conditional testing mechanism works as a simple and efficient inferential rule. "Testing" simply means that each "case" (persons-situation-consequences) is evaluated to the extent that it does or does not express conformity to the command.

However, the conditional rule (IV) should not be reduced to the command: on the contrary, the command is one of the parts of the rule's system or checking mechanism. In other words: a more abstract rule encompasses more particular rules - in turn, the checking mechanism is sensitive to the belief system and to context variation.

(V). *Categorial construal*: Finally, the concept used to label the case is usually embedded within a categorization judgment. For instance: "The action should not be carried out because it violates the (unconditional) command, therefore, it is a *wrong* action." Note that the adjective *wrong* does not have any sensory-based feature that could be classified as a "wrong" attribute (Cf. Barsalou, 1987, 2017; Stich, 1993).

The command is the *reason* why the action should not be done. In turn, the term used to categorize the case (in this case, "wrong") derives from the fact that the checking mechanism has verified the violation of the command. Accordingly, the (CTC) model will predict that any event of moral conceptualization will refer the case directly to a command, and only afterwards it will be possible to use a specific category to label the case - always using the command as the standard of comparison. In other words, moral conceptualization and generalization makes sense only to the extent that the observational data is compared to a norm that has internal validity regardless of the configuration of the data, and of the subjective preferences of the parties involved.

The descriptive stage (I) is a statement of mere empirical facts. The level of non-normative assessment (II) shows that empirical facts can have more than one reading: there may be contrary opinions about how to *frame* the case itself. In the hypothetical case exposed, the driver offers a non-normative evaluative framing as its own "version" of it.

The normative-question level (III), on the other hand, puts all the emphasis on the applicability of impersonal and unconditional duties. Note that when we enter the realm of unconditional duty, it is not possible to reduce statement (III) to statements (I) and (II) - these are not interchangeable expressions.

The categorization of the act following normative assumptions depends on the answer given to question (III). However, what "should be done" is also often the subject of different perspectives. The more specific the command is – e.g., the sign that prohibits throwing garbage into the river - the less indefinite its application becomes, but at the same time, the more restricted its scope. Kant considers that this cognitive phenomenon - known in logic as the "rule of inverse variation of extension and intension" - is the fundamental characteristics of the human conceptual system (Kant, 1992).

5. Mutual consent, contractual obligation, and moral obligation

This brings us back to the question of how commands acquire a special sense of "moral" oughtness. At this point let me use as an example the social contract theory in its classical version, in the tradition that goes from Engelbert of Admont to John Locke, which understand the "social contract" as a theoretical construct that is not merely hypothetical but also based on certain historical and social processes (Cf. Lessnoff, 1986).

Now, suppose we have access to an imaginary defender of classical social contract theory. This social contract theorist believes that all the known rules, both conditional and unconditional, have their origin in a long-term process of cumulative social agreements that constitute some sort of social pact or "contract". Looking from this perspective, he/she suggests that the specific sense of "moral" oughtness, unlike the sense of a pre-moral ought, may be the result of a question such as:

"Why don't we take some possible patterns of behavior whose application is usually conditional on the actual behavior of others – under logics such as the quid pro quo, and the "tit for tat" - and consider them as if they were unconditional duties that apply equally to everyone, and to any possible conflict of interests, regardless of partial wishes and goals, also regardless of what the other does, and without expecting any reward for doing and respecting them?"

Once this proposition is accepted, certain actions will be performed, and others will be avoided, not in relation to certain partial interests, but on the contrary, because one performs them for goodness and justice's sake, which implies that one complies with the principle for the sake of the principle, therefore, not obeying contingent conditions alien to the principle itself. It is a type of attitude that could be summed up in a sentence such as: "I respect you, or them, for the sake of acting under the principle of respect for all people - no strings attached".

However, we might object to this idea in this way:

"That may be fine for the origin of certain specific duties usually called as "natural" or "moral," but it does not explain the origin of the sense of duty itself, that is, of the sort of conditional duty that is prior to the unconditional formula of "acting for goodness/justice's sake".

To which our imaginary theorist of classical social contract theory might reply:

"Fair enough, the initial conditional "ought" that serves as the basis for building ulterior unconditional duties cannot in turn derive from another ought - at the risk of incurring in an infinite regress. To stay within the bounds of social contract theory, I would understand the premoral "ought" as the product of everything that the members of a group decide to establish as a binding promise or pledge.

In a purely contractual sense, I produce a duty when I make assurances that I will fulfill my part of some agreement - on the basis that others will fulfill theirs. The first notion of an "ought" derives in this case from a reciprocal promise of future performance between the parties to an agreement. No need to assume that there are obligations prior to this initial agreement."

However, the discussion does not end here. The hypothetical position in which we have placed our "social contract theorist" is interesting precisely because it is falsifiable, if viewed from another point of view. For example, David Hume could still argue that it is not the promise that produces the sense of duty, but rather the sense of duty that produces the promise – leaving aside the obvious fact that there are mutual promises based on the self-interested motives of each of the parties that do not depend upon a previous "sense of duty". And where does this so-called "sense of duty" come from?

According to Hume, this comes from our passions and natural inclinations - therefore, "where an action is not required by any natural passion, it cannot be required by any natural obligation" (Hume, 2006, p. 112). Seen from this angle, moral oughtness depends on natural passions that work for the common good, and *only* those natural passions are the origin of moral obligations, even in the absence of the reciprocal consent presupposed by social contract theorists (presumably Hume is criticizing Hobbes).

Faced with this Humean argument, the defender of the contractual position could still reply that:

"The sense of oughtness does not depend on passions, which usually act in a reckless and violent way on the spirit, but instead on the free will of the parties. Contracts create an obligation by the mere act of mutual consent. Consent, when it is dictated by conscience, and is given willingly, without submitting to any imposition or intimidation, gives rise to obligation."

The discussion does not have to end here either. Hume could still argue that the contracting parties have an obligation to act in good faith – e.g., as exemplified in the legal concept of an "implied covenant of good faith and fair dealing", article 1–304 of the "Uniform Commercial Code" of the United States - with which there would be at least *one* type of obligation that is prior to the agreement between the contracting parties (and it is precisely a "moral" kind of obligation).

To counter this example, the defender of contractarianism - understanding by "contractarianism" both the moral theory, and the political theory (for the distinction see Cudd & Eftekhari, 2021) - could still argue that the commitment to act in good faith is materialized once both parties have given their reciprocal consent. In other words, good faith can only be promised when the other party is equally committed to it, therefore, more than an antecedent, it is a product of the reciprocal consent. This would be the most rational way to prevent one of the parties, the one who would want to benefit by acting in bad faith, from taking advantage of, deceiving, or exploiting the other, the one who, instead, would act in good faith.

This hypothetical debate between our two thinkers about what is first and what is second regarding consent and obligation can be endless. Let us admit for a moment the point of view of the advocate of moral contractarianism. What implications does that position have for the matter at hand?

From this specific point of view, the moral ought would be the product of two main steps: first, a mutual consent that gives rise to a premoral obligation - since mutuality is not "moral" *per se*: not all mutual agreements are just or good for all parties - then, a question, but also an attitude, or a stance, that takes certain conditionally binding clauses and transforms them into an unconditional formula, which demands allegiance to the principle of acting for goodness and justice's sake.

If this approach is correct, then any theory that remains exclusively anchored to explanations of the first stage - i.e., the stage of mutual agreements and reciprocal contributions - could hardly be a truly *explanatory* theory of morality. Thus, for example, it is not simply a question of contractually defining a specific wage, and there are several economic and sociological theories about wage-determination, but also - and mainly - of defining a *just* standard of remuneration according to each situational context (Stabile, 2016).

In a similar way, certain types of contracts include some clauses that protect the weaker parties in those situations where there is a manifest asymmetry of power - that is, situations in which the most powerful party can dictate (impose) at will the terms of exchange to the weaker party. Protecting the weakest means putting into play the principle of acting for the benefit of what is good and fair - otherwise it would be a clear case of abuse. I surmise that it is at this second stage that we can really speak of an ought with a "moral" sense.

In summary, moral obligations are *superior* to contractual-conditional relationships between individuals, or between individuals and organizations (a point on which Hume would agree). This assumption is extremely important and has legal consequences: just consider much of the juridical advances of the last years in the United States, Canada, and Europe regarding social justice issues. For instance, the different laws and policies condemning and prosecuting acts of violence against women, the legislation abolishing human trafficking, or the regulations enacting that even if a work contract is illegal, the employer has certain responsibilities regarding the basic protection of its employees. All these regulations are based on the idea that some basic rights and duties must be unconditionally upheld.

6. Conceptual structure

The principle of acting for goodness and justice's sake, as well as any other moral principle, always occurs within a conceptual structure.

Ray Jackendoff defines the "conceptual structure" as "a means for identifying and categorizing [tokens]: the rules specify ways of checking input against an internal standard" (Jackendoff, 1999, p. 140). In turn, he points out that these rules can be of two types: either a *system of necessary and sufficient conditions*, in which the case must satisfactorily comply with all the conditions established in the category, or rather a *preference rule system*, in which certain prototypical images are used as default values, therefore, one "does not have to check all the conditions to arrive at a judgment" (Idem, 140–142).

For both theories, a category is a collection point that connects the different features that are attributed to a kind (Murphy, 2002, p. 309).

These two models presuppose that "by knowing the category to which a thing belongs, the organism, therefore, knows as many attributes of the thing as possible" (Rosch, 1975, p. 197). However, while the first model only admits perfect correlations between attributes within a category, the second approach, based on prototypical categories, "allows membership to entities which share only few attributes with the more central members" (Taylor, 1995, p. 54).

In any case, the checking mechanism proposed by (CTC) is not limited to either of the two categorization models described above. Perhaps the conceptual architecture of semantic memory is closer to the second model than to the first. In this case, the conceptual system would be mostly informed by fuzzy sets, that are organized around a basic level of categorization - something that Aristotle himself recognizes, when he points out that if someone wants to say what the primary substance is, it is more explanatory to indicate what their species is than what their genus is (See chapter V, *Categories*). However, even within a general fuzzy system there is nothing to prevent certain well defined and clear-cut conceptual representations from arising.

For the case at hand, the central question is, as McHugh et al. (2022) suggested, to determine what are the features that we can attribute to the very notion of "category".

For instance, if the content of a concept is its "referent", plus the "compositional structure of the mental representation that expresses the concept", as Fodor and Pylyshyn (2015, p. 130) pointed out, it is difficult to establish what could be the typical "referent" of an abstract category such as justice - and certainly it is not reduced to some conventional allegory. In the same way, it is difficult to find prototypes for very abstract ideas such as prudence, wisdom, or the sense of humanity.

It seems that the most general concepts either have the largest possible number of exemplars (e.g., the addition of “human 1” plus “human 2”, and so subsequently), and/or they have a very restricted set of basic features (e.g., “human”: living being + primate + talking primate). What they cannot do is to tie themselves to a typical exemplar or a prototypical member (such as: “"talking primate that has such a color, height, religion, etc."). For instance, in an imaginary context of widespread public risk, caused by governmental or corporate malpractice, and which affects all social sectors equally, what could be the “typical member” of an expression such as “patient who suffers an act”?

The relevant issue is that the most general categories are those that allow a greater degree of *variation* - something that works in favor of exemplar theory, but not of prototype theory, as Murphy (2002, pp. 41–42) acknowledges. In turn, the problem with exemplar theory is that it leaves out all those novel cases of categorization that do not have a similar empirical precedent. The most general categories have a problematic relationship with the logic of similarity between case and concept basically because the *more* abstract the concept is, the *fewer* features it has, thus fewer opportunities for comparison. Consequently, the categories with the higher degree of generality - like those used in logic - are simply featureless.

Instead, the CTC model simply assumes that, for the case of moral evaluation, first, there is a cognitive verification of certain premises (generic and unconditional ideals), then, depending on the degree of application of those premises, certain categories are activated, usually embedded within moral judgments. In no case should we assume that the conditional rule is something that works "inside" the categories. It is not a matter of comparing features of the case with features of the category – more than these types of similarity judgments, categorization and category learning seems to be “special cases of inference to the best explanation" (Rips, 1989, p. 53).

But I surmise that the issue also goes beyond a mere heuristic procedure. The point is that if you have access to a command –employed as standard of comparison - whose structure is not reduced either to the case or to the category (label) used to represent the case, as I have shown in section 4, the entire categorization process is no longer dependent on the category’s internal configuration. Lexical concepts become pure pragmatic (cultural) conventions.

Furthermore, this idea operates in the field of syntax as well: the set of possible kinds that are constructed grammatically do not depend only on the "internal" structure of the categories, but also on other morphosyntactic constituents of the propositions - think, for example, of the way in which language classify different types of actions through prepositions, definite articles, partitive articles, qualifying adjectives, adjectives of appreciation, pronouns, among other forms that go beyond the nouns themselves.

The main difference between the categorization models mentioned above and the theory proposed here therefore resides in the definition of the standard or reference point used for moral categorization, generalization, and judgment. While both the “model of necessary and sufficient conditions” and the “preference rule system” - to use Jackendoff’s terminology - consist of comparing the features of A (case, referent) with the features of B (concept), where B is used as the standard of comparison between similar cases, the model of “categorization by testing the command” (CTC) contains not two, but rather three basic elements, namely: A (case), B (concept), and C (command), where the standard of judgment and comparison is no longer B, but C.

In short, the *moral* evaluation system accepts all those behavioral variations that, by satisfying the demands of the commands, or at least by not contradicting them, are rendered permissible.

7. A system of semantic polarity

One of the peculiarities of the lexical concepts typically used in moral evaluation is that these seem to be immersed in relations of logical and semantic antonymity (Cf. Allan, 1986; Osgood & Sebeok, 1965). It implies that certain conceptual representations – both entries in the mental lexicon, and inferentially derived concepts - are usually processed, codified, and interpreted as belonging to a chain of *opposite* semantic poles, in which the positive side (e.g., the behavior that is ethically desirable and mandated) is represented as opposed to the negative side (e.g., the behavior rendered to be ethically impermissible and prohibited).

The term "opposite", used here only in a conceptual sense, without ontological pretentions, means that the function of each sense unit depends on its relation of antagonism with a contrary sense unit (Croft & Cruse, 2004, p. 112). Taking as an example two opposite points, a positive pole versus a negative pole (its counterpart), the value of each of them depends on the inversion of the value of its opposite (Saussure, 1997, p. 8a). This system of semantic polarity between different cognitive types is based on relations of contradiction, mutual denial, antinomy, and mutual exclusion. From a logical point of view, abstracting from any empirical situation, there is the highly idealized expectation that opposite categories will reflect a relationship of mutually exclusive codetermination.

There is abundant literature on the processes of semantic opposition that are required in the construction of evaluative judgments. This tradition goes from Edward Sapir’s studies on the judgments of grading in which all comparative terms are measured from a standard that is shared by both opposite poles - for instance, good is better than indifferent, whereas bad is worse than indifferent (Sapir, 2008, p. 449); going through John Lyons’ suggestion that terms like good and bad can be either contradictory, in which case they negate each other, or simply contrary, in which case the negation of one does not logically imply the assertion of the other (Lyons, 1968, p. 461); until arriving at the research made by Magnus Ljung on how we evaluate "natural" concepts such as "king", "painter", "mother", and "typewriter" in terms of how well they express the ideas of goodness and badness (Ljung, 1974, p. 79); and the studies carried out by Jerrold Katz, who analyzes how evaluation semantic markers like “good” and “bad” function as lexical readings of a specific type of nouns in predicate adjective sentences (Katz, 1964, p. 757).

However, it is important to note that it is not necessary for the two opposite terms to be grammatically marked. In some natural languages only one of the two opposite concepts is overtly part of the vocabulary. For example, in Xhosa, the adjective class contemplated a term for "bad/ugly" but lacked a term for its opposite concept "good" (McLaren, 1936, p. 63). In this example, both concepts presuppose each other, although only one of them is instantiated in the list of adjectival forms. Here, as in so many other cases studied by structural linguistics, what is marked ("bad") is what deviates from the norm ("good"). It is not necessary to mark the norm since it is obvious to everyone.

The contrasting poles are typically conveyed by antonym terms. For instance,

(a) The polarity between the contrary notions of inviolability vs. violability, materialized in the recognition of certain inherent natural rights and duties versus their denial. The semantic pole of inviolability is derived from the ethical ideal that human beings have an inherent dignity, linked to the prescription that we should not treat people as if they were mere instruments. That is to say, the dignity of one person should not be subordinated to the interests of another person(s) - aka the Kantian "formula of the end in itself" (Dworkin, 2011).

In turn, the belief according to which people are inviolable and equally important is linked to the prescription that any kind of harm must be prohibited, and that every person should be equally and impartially treated, which brings us to,

(b) The semantic pair of impartiality vs. partiality. These opposites poles are involved in a competitive relationship in which the winning pole (i.e., the one chosen or applied), displaces the contrary, thereby cancelling all its semantic content, or using it in the opposite way (Saussure, 1997, p. 11a).

Although the moral point of view is usually associated with the notion of "impartiality", it is also possible to establish a contrast between the moral and non-moral dimensions using the concept of "partiality", as in,

c) Benevolent partiality vs. malevolent partiality. That is, a partial attitude can be rendered as "moral", if it is labeled as benevolent (Cf. Paolacci & Yalcin, 2020). In certain cases, the principle of "benevolence" prevails, much more than that of "neutrality". It makes perfect sense, since the principle of benevolence includes some of the basic features of the moral point of view – e.g., generic reference, the ideal of solidarity, and the principle of personal responsibility.

As a matter of fact, the concepts we use in moral evaluations - such as the opposites *partiality* versus *impartiality* - are not "moral" or "immoral" in themselves, because every concept is subject to syntactic and semantic ambiguity. From a purely semantic and pragmatic standpoint it is difficult to accept the idea proposed by the defenders of "moral foundations theory" (MFT) according to which moral cognition is implemented by sets of related modules working together to solve specific problems (Graham et al., 2013). There is no empirical evidence that allows us to even suppose that there are certain mental "modules" specialized in processing the specific *moral* sense of a category, versus other modules in charge of processing the *non-moral* sense of the same category.

The crucial point is that normative ideals, understood as the mental representations of the type of behavior that is both desirable and responsible, always stand on the side of one of the semantic poles, the pole of the virtues or the superior and inalienable values expressing the morally good and right (e.g., the *summum bonum,* the *fair mean,* the *golden rule*), and, therefore, one of its main features is the property of exclusivity (i.e., exclusion of the opposite semantic pole). The opposite pole does not have to be a single concept, but can be a plurality of semantic notions, an important issue highlighted by Aristotle (2014, p. 249-251).

The moral point of view is just a criterion, based upon normative ideals, for discerning and choosing between competing beliefs, motivations, reasons, and actions. There may be the temptation to link the normative ideals with the prototypes studied by Eleanor Rosch, or with "cognitive templates" that are built "from" the features of the exemplars (Gray, Young, & Waytz, 2012, p. 102) - but these options are fraught with certain logical and theoretical problems.

Both feature-based templates and prototypes are anchored to specific attributes, grounded on precise phenomenological data (Cf. Murphy, 2002), while normative ideals, on the other hand, do not depend on specific basic-level attributes. As Paul Ziff pointed out, referring to the difficulty of reducing the meaning of the adjective "good" to the supposed phenomenological "properties" of something: "If we weld something's being good to its having certain characteristics, certain specific characteristics in each case, then the word "good" is likely to be replaced by something like "Grade A″ or "premium Grade" "(Ziff, 1960). This problematic issue does not preclude that some prototypes, understood as stereotypical members of a category, can be used as the cultural models of correct behavior, but in any case, normative ideals, and ethical beliefs in general should not be reduced to being mere prototypes – e.g., some cases that have no similarity relation to the "prototype", not even a distant relation, can also be classified as occurrences of the same type.

8. Semantic variation in moral cognition

Every command conveying an impersonal-unconditional obligation is the expression of an ideal of conduct. However, one might ask: Why do we need commands and normative ideals to help in the process of "moral" categorization, if we can make use of evaluative lexical concepts already stored in the lexicon?

The problem is that to decide if a concept has a specific "moral" significance, it is necessary to compare it with a standard - i.e., with a mandate that expresses an impersonal and unconditional duty not reducible to individual interests - that goes beyond the properties attributed to the concept itself. Furthermore, we need to resort to the sense of unconditional obligation (explored in section 5) because certain lexical concepts typically employed in moral judgment, such as good, better, or true, can also be used in non-moral and immoral ways – e.g., people can be "very good" at deceiving or hurting others (*Lesser Hippias*, 371e).

It seems that there is no superordinate category that includes at the same time the sense of "stealing is not *good*" with the opposite sense of "that's a *good* stealing". A similar difficulty emerges with the adjective "right", which can mean righteousness, in a moral sense, but can also allude to the right side in spatial terms, in a non-moral sense. It seems that both the moral point of view and the non-moral point of view make use of the same delimited set of categories (the classic problem of *penuria nominum*). The context is not enough to solve this problem because the context itself - be it syntactic, semantic, or pragmatic - can be ambiguous.

The same problem of semantic variation applies to those theories of moral categorization that are based on a procedure that compares the attributes of a case with the attributes of a concept. The authors supporting this argument simply overlook that phenomenological similarity between tokens does not necessarily imply that they share the same type (homonymy), while observable differences between tokens can perfectly translate into a common meaning (synonymy). For instance, two different emotional reactions can be an expression of the same type, or vice versa: similar passions can relate to entirely different concepts. Therefore, it is difficult to accept the idea that moral conceptualization is based on a direct matching between case and concept. At the very least, the data alone does not allow us to decide which is the correct evaluative interpretation - an example of the classic Duhem-Quine's thesis of underdetermination of theory by data.

Trying to explain the mechanism of moral conceptualization, Jesse J.

Prinz wonders what "perceptual features" could be involved in the processing of abstract concepts. His answer is that we learn the meaning of abstract notions such as "good" and "bad" through a kind of "inner-perception" whereby we "observe" how we react to certain situations, and then construct labels to represent those emotional reactions. In other words: the body gives us the adequate physiological response, and this sensory-based and motor answer is the basis for the ulterior abstract conceptualization (Prinz, 2005).

However, Prinz overlooks the fact that concepts like "good" and "bad", "right" and "wrong", are not monosemic concepts, but rather are subject to semantic contrast, variation, and ambiguity, ranging from vagueness and synonymity, to polysemy and homonymy. Even if the galvanic response were always the same, people may be using these concepts to convey *different* meanings and connotations. That is the main problem that derives from separating conceptual knowledge from commands, and vice versa: that one may come to suppose that the "moral" meaning is primarily given by the phenomenological (i.e., perceptual/motor) "features" that one attributes to a concept. But we cannot simply assume that there is a clear *correspondence* of attributes between type and token, or that both logical levels share a *single* meaning (monosemy), which is supposedly condensed into a *single* concept.

To complicate matters further, categories such as inviolability (dignity) and fairness (just treatment) do not contain *per se* a particular clause stipulating that these notions are to be understood as "moral" mandates even if they are expressed as simple suggestions. In the absence of an unconditional command, a large part of the evaluative concepts that range from the *best* to the *worst* could not be invested with a proper moral meaning.

Accordingly, I surmise one of the core functions of the commands is to help the connection between certain categories within the semantic system with the modal norms of obligatoriness and permission. The *normative* structure of moral conceptualization is perfectly explainable without the need to assume specific "moral" contents – in line with Bucciarelli, Khemlani, and Johnson-Laird (2008), Sinnott-Armstrong (2012); Barsalou (2017), and McHugh et al. (2022) - but instead it is essential to know what is the duty that serves as the ultimate criterion for ethical evaluation and judgment.

In short, the process of moral categorization and generalization consists of using a series of evaluative semantic markers - such as *right* and *wrong*, among many others - within judgments concerning what should be done, as opposed to what one could (or would like to) do but should not.

As "right" and "wrong" are eminently ambiguous concepts, conveying both non-normative and normative facets, moral judgments cannot be the result of a mere pairing between features of the case and features of the concept. Both levels are prey to semantic variation, to say the least. Therefore, we can only categorize in a moral sense from the moment in which actions, omissions and consequences are semantically covered by a typology of acts of impersonal-unconditional duty, embodied by specific commands in all its different formats, from the polite suggestion to the categorical imperative.

9. Conclusion

The hypothesis that I have offered is the following: the act of moral categorization does not consist of a judgment of similarity between case and concept, but rather it is based on a cognitive operation of comparison between case and command, carried out by a checking mechanism. The checking mechanism is a conditional rule that measures the degree of compliance of a case with a command. Conformity of the case with the command activates a specific type of concepts, while non-conformity activates the opposite concepts.

Why is it problematic to assume that moral categorization consists of a simple pairing between cases and concepts? There are some important reasons for this, among them,

First, the set of lexical concepts typically used to express a moral evaluation have no iconic or phenomenological similarity to the cases they represent,

Second, both the terms we use to *describe* the case, and the concepts used to *categorize* the case, are subject to a wide degree of semantic variation, rendering inoperative any claim to a mere pairing between type and token that is not mediated by some disambiguation factor.

The disambiguation factor that I have proposed is the notion of a command that expresses an impersonal-unconditional obligation (and it expresses it in all possible ways, including the use of conditional constructions). Sentential and pragmatic contexts cannot serve as disambiguating factors because contexts themselves may be ambiguous unless normative constraints are included.

As a result, even if there is no apparent analogy between different cases, they may be represented by the same category. By the same rule, cases that are perceptually similar can be categorized differently. The fundamental issue is that abstract principles are not isomorphic to observable properties. In moral cognition what is being compared is an action with a command, plain and simple. Commands carry a normative meaning that no single word can exhaust, and that is not reducible to any specific category – among other reasons, because lexical concepts are not monosemic. The core meaning of commands is derived from the set of unconditional duties and ideals, and then it is adjusted to each type of possible action and consequence.

The task of addressing the "normativity problem" exposed in this article – that is, the problem of how to connect cases with moral ideals - depends on a very practical interrogation, namely: Are the commands actualized or not, and to what extent are they actualized? Note that, by default, I have assumed that the actualization of the commands, even in their best examples, is more approximate than absolute. Despite statistical regularities, each concrete case is different from all the others: it builds a specific way of approaching the set of practical norms, and those ways are not necessarily identical to each other.

10. Appendix: Glossary of the basic terms (constituents) of the model. See Section 3

Case: mental representation of some specific situation/actors/consequences, including both observable and non-observable properties.

Command: mandate ordering what should and shouldn't be done, therefore involving the expectations of compliance and execution.

Concept: logical type that groups a series of objects or events within certain specific lexical meanings.

Conditional rule: Cognitive mechanism that checks if certain commands are satisfied or not in specific cases, and to what degree they are satisfied, from which a categorization judgment is derived.

Author statement

The corresponding author is responsible for ensuring that the descriptions are accurate and agreed by all authors: The article has a *single author*, who has uniquely contributed to the following items:

Term: Definition.

Conceptualization: Ideas; formulation or evolution of overarching research goals and aims.

Methodology: Creation of the theoretical model.

Software: No software involved.

Validation: No experiment involved - it is a purely theoretical model.

Formal analysis: No formal analysis involved.

Investigation: Conducting the analysis process.

Resources: analysis tools are purely theoretical - no access to specific data sources.

Data Curation: Does not apply.

Writing - Original Draft: Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)

Writing-Review & Editing: Preparation, creation and/or presentation of the published work by the author, specifically critical review, commentary or revision – including pre-or postpublication stages.

Visualization: Preparation, creation and/or presentation of the published work.

Supervision: Oversight and leadership responsibility for the research activity planning and execution: single author.

Project administration: the author.

Funding acquisition: No funding.

Data availability

No data was used for the research described in the article.

References

Allan, K. (1986). *Linguistic meaning*. New York: Routledge.

Aristotle. (2014). *Aristotle's ethics: Writings from the complete works*. In Jonathan Barnes, & Anthony Kenny (Eds.). Princeton, NJ: Princeton University Press.

Atran, S. (2021). The will to fight. *Science, 373*(6559), 1063-1063.

Barsalou, L. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge University Press.

Barsalou, L. W. (2017). Cognitively plausible theories of concept composition. In J. A. Hampton, & Y. Winter (Eds.), *Compositionality and concepts in linguistics and psychology* (pp. 9–30). Springer.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating Abstract Concepts. In D. Pecher, & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 129–163). Cambridge University Press.

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2016). Moral judgment and decision making. In G. Keren, & G. Wu (Eds.), *The wiley blackwell handbook of judgment and decision making* (pp. 478–515). Chichester, UK: Wiley.

Bucciarelli, M., Khemlani, S., & Johnson-Laird, P. N. (2008). The psychology of moral reasoning. *Judgment and Decision Making, 3*, 121–139.

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology, 17*, 391–416.

Clancy, P. M., Akatsuka, N., & Strauss, S. (1997). Deontic modality and conditionality in discourse: A cross-linguistic study of adult speech to young children. In A. Kamio (Ed.), *Directions in functional linguistics* (pp. 19–57). Amsterdam: John Benjamins.

Cosmides, L., Guzmán, R., & Tooby, J. (2018). The evolution of moral cognition. In A. Zimmerman, K. Jones, & M. Timmons (Eds.), *The routledge handbook of moral epistemology* (pp. 174–228). New York: Routledge Publishing.

Croft, W., & Cruse, A. A. (2004). *Cognitive linguistics*. New York: Cambridge University Press.

Dworkin, R. (2011). *Justice for hedgehogs*. Cambridge, MA: The Belknap Press.

Fodor, J., & Pylyshyn, Z. (2015). *Minds without meanings: An essay on the content of concepts*. CA, Massachusetts: The MIT Press.

Geertz, C. (1973). *The interpretation of Cultures: Selected essays*. New York: Basic.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., et al. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine, & A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 47, pp. 55–130). Academic Press.

Gray, K. J., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124.

Greene, J. D. (2015). The rise of moral cognition. *Cognition, 135*, 39–42.

Grice, P. (1991). *Studies in the way of words*. Cambridge, MA: Harvard University Press.

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, pp. 367–391). New York: Oxford University Press.

Harman, G., Mason, K., & Sinnott-Armstrong, W. (2010). Moral reasoning. In J. M. Doris (Ed.), *The moral psychology handbook* (pp. 206–245). Oxford University Press.

Hauser, M. (2006). *Moral minds. How nature designed our universal sense of right and wrong*. New York: Harper Collins Publishers.

Holyoak, K. J., & Cheng, P. W. (1995). Pragmatic reasoning with a point of view. *Thinking & Reasoning, 1*(4), 289–313.

Hume, D. (2006). Moral philosophy. In *With introduction by geoffrey sayre-McCord*. Indianapolis: Hackett.

Jackendoff, R. (1999). *Semantics and cognition*. Cambridge, MA: The MIT Press.

James, A. (2017). The moral continuum: Congruence, consistency, and continuity in moral cognition. *Theory & Psychology, 27*(5), 643–662.

Johnson, M. (2014). *Morality for humans: Ethical understanding from the perspective of cognitive science*. Chicago: Chicago University Press.

Kant, I. (1992). Lectures on logic. In *J. Michael Young*. New York: Cambridge University Press. and translated by.

Katz, J. J. (1964). Semantic theory and the meaning of "good". *The Journal of Philosophy, 61*(N. 23), 739–766.

Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory in Practice, 16*(N. 2), 53–59.

Kohn, M. L., & Schooler, C. (1983). *Work and personality: An inquiry into the impact of social stratification. Norwood*, N.J: Ablex Publishing.

Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are based on common principles. *Psychological Review, 118*(1), 97–109.

Kruschke, J. K. (1992). Alcove: An exemplar-based connectionist model of category learning · John K. *Psychological Review, 99*(1), 22–44.

Lessnoff, M. (1986). *Social contract*. London: Macmillan.

Liao, S. M. (Ed.). (2016). *Moral brains: The neuroscience of morality*. New York: Oxford University Press.

Ljung, M. (1974). Some remarks on antonymy. *Language, 50*(N. 1), 74–88.

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge: University Press.

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2022). Moral judgment as categorization (MJAC). *Perspectives on Psychological Science, 17*(1), 131–152.

McLaren, J. (1936). *A Xhosa grammar*. London: Longmans.

Mikhail, J. M. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. New York: Cambridge University Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: The MIT Press.

Osgood, C. E., & Sebeok, T. A. (1965). *Psycholinguistics: A survey of theory and research problems*. Bloomington: Indiana University Press.

Paolacci, G., & Yalcin, G. (2020). Fewer but poorer: Benevolent partiality in prosocial preferences. *Judgment and Decision Making, 15*(2), 173–181.

Prinz, J. J. (2005). Passionate thoughts: The emotional embodiment of moral concepts. In D. Pecher, & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 93–114). Cambridge University Press.

Rawls, J. (1999). *A Theory of justice* (Revised edition). Cambridge, Massachusetts: Belknap Press.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou, & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

Rokeach, M. (2008). *Understanding human values*. New York: Simon and Schuster.

Rosch, E. (1975). Universals and culture specifics in human categorization. In R. W. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning* (pp. 177–206). New York: John Wiley and Sons, Sage Publications.

Sapir, E. (2008). *The collected works of edward Sapir I: General linguistics*. New York: Mouton.

Saussure, F de (1997). *Deuxieme Cours de Linguistique Générale (1908-1909) d'après les cahiers d'Albert Riedlinger et Charles Patois*. New York: Elsevier.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology, 25*, 1–65. Elsevier.

Sinnott-Armstrong, W. (2012). Does morality have an essence? *Psychological Inquiry, 23*(2), 194–197.

Stabile, D. R. (2016). *The political economy of a living wage: Progressives, the new deal, and social justice*. St. Mary's City, Maryland: Palgrave.

Stich, S. (1993). Moral philosophy and mental representation. In M. Hechter, L. Nadel, & R. E. Michod (Eds.), *The Origin of values* (pp. 215–228). A. de Gruyter.

Taylor, J. R. (1995). *Linguistic categorization: Prototypes in linguistic theory*. New York: Oxford University Press.

Tetlock, P. E. (1986). A value pluralism model of ideological reasoning. *Journal of Personality and Social Psychology, 50*, 819–827.

Ziff, P. (1960). *Semantic analysis*. New York: Cornell University.