

THE REPRESENTATIONAL FORMAT OF MORAL CONCEPTS FOR SITUATED-ACTION

by

Uphaar Dooling

Copyright © Uphaar Dooling 2021

A Thesis Submitted to the Faculty of the

DEPARTMENT OF PHILOSOPHY

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2021

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master’s Committee, we certify that we have read the thesis prepared by: **Uphaar Dooling**
titled: **The Representational Format of Moral Concepts for Situated-Action**

and recommend that it be accepted as fulfilling the thesis requirement for the Master’s Degree.



Sara Aronowitz

Date: Aug 15, 2021



Peter Jansen


Date: Aug 15, 2021



Jordan Theriault

Date: Aug 15, 2021

Final approval and acceptance of this thesis is contingent upon the candidate’s submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master’s requirement. 



Sara Aronowitz
Thesis committee chair
Philosophy

Date: Aug 15, 2021



Table of Contents

Abstract.....	4
Section 1: The Features of Moral Concepts.....	8
Section 1.1: The Descriptive and Evaluative Components of Moral Concepts.....	10
Section 1.2: Moral Concepts and Moral Motivation.....	13
Section 2: The Invariantism-Contextualism Dichotomy as Candidates.....	14
Section 2.1: Clarifying the Invariantism-Contextualism Debate.....	16
Section 2.2: Traditional Psychological Theories of Concepts.....	19
Section 2.3: Accommodating Moral Concepts in Traditional Psychological Theories of Concepts.....	23
Section 3: The Challenge of Moral Experience and Situated-Action for Traditional Psychological Theories of Moral Concepts.....	31
Section 3.1: Introducing a Moral Vignette.....	32
Section 3.2: Moral Concepts Possess Variable Evaluative Content and Motivational Force.....	35
Section 4: Context-dependent Representational Format and Ad Hoc Concepts.....	43
Section 4.1: Introducing Ad Hoc Concepts.....	44
Section 4.2: Ad Hoc Moral Concepts.....	47

Conclusion.....51

References.....52

Abstract

Moral concepts serve central roles in facilitating our everyday moral thought and behavior. Despite their importance, few attempts have been made to provide or characterize a satisfactory psychological theory of moral concepts that can accommodate their representational features and explain their function within our everyday moral experiences. In this thesis, I argue traditional psychological theories of concepts that posit invariant representational formats cannot provide a satisfactory theory of moral concepts while accommodating their motivational roles in the selection of situationally appropriate and time-sensitive action. The evaluative content and motivational force of a tokened moral concept varies across each instance and is highly situated to the agent's present environment, cognitive tasks, constraints of the situation, etc. Invariant representational formats of concepts are unable to accommodate for the variable conceptual content and are unfit to ground a theory of moral concepts while maintaining their central roles within facilitating our moral cognitive capacities. Instead, I argue, any viable psychological theory of moral concepts must posit a flexible and context-sensitive representational format to meet the variation present during instances of moral experience.

Keywords: Concepts; moral concepts; contextualism; invariantism; ad hoc concepts

The Representational Format of Moral Concepts for Motivating Situated-Action

Imagine visiting a familiar coffee shop while on your way to work. As you enter the building, you recognize a long line of customers waiting to place their orders. Although in a hurry, you make a conscious decision to endure the wait as you navigate from the entrance to the end of the line. While waiting for your turn to order, you mindlessly survey the building and people inside until your attention is drawn to a specific event – a hurried customer drops their wallet while making their way towards the exit. Immediately, you feel an internal motivational force compelling you to act, and without explicit conscious planning, you walk towards the customer, pick up the wallet, tap their shoulder, and extend your arm to return their valuable possession.

Scenarios, like the one above, are not uncommon in our everyday lives. As we navigate through our social environments, numerous situations call on us to think and behave morally. Fundamental to the various neurocognitive processes at work is the ability to represent information by generalizing quickly, accurately, and reliably from our past experiences and learned associations (Bloch-Mullins, 2015). Representational approaches to cognitive science argue *concepts* facilitate this ability (Barsalou, 2012; Bloch-Mullins, 2015; Laurence & Margolis, 1999; Löhr, 2017; Machery, 2009). Concepts are commonly identified as mental representations (Margolis & Laurence, 2007) that are neuronally instantiated as mental states representing features, properties, objects, or other information about some state of affairs in the world (Barsalou, 2012; Machery, 2009; Michel, 2020). Concepts allow us to pull from our previous experiences in similar situations to categorize our current sensory information into meaningful bundles that can be productively used by higher-order neurocognitive processes (e.g., the construction of inferences, issuing of motor commands, production of language, etc.). In our coffee shop scenario, the representation of information corresponding to the customer's wallet is

achieved through the goal-directed and generative process of interpreting some bundle of salient sensory information. Our ability to mentally represent the content of the concept **WALLET** stands-in for that bundle of information and confers meaning to this process of identification and representation.¹

During instances of moral significance, moral concepts – e.g., **FAIRNESS**, **RIGHTNESS**, **HARM**, **WELLBEING**, etc. – are used to represent morally relevant sensory information. We token the concept of **WELLBEING**, in the above example, as we recognize the potential harm associated with losing a valuable possession. An interesting feature of moral concepts is that their tokenization will attribute moral value or significance to the represented information by imposing a normative evaluation onto some state of affairs in the world (Fingerhut & Prinz, 2018). This evaluation confers to moral concepts their motivational force allowing them to facilitate the formation of moral judgments and the selection of action. However, no two moral experiences are the same – each experience requires representing moral concepts that are sensitive to our past experience, our current goals or intentions, and situational constraints. Not all instances of representing the concept **WELLBEING** will motivate us to perform the same actions, or even evaluate its content in the same manner. A central challenge for moral cognitive science is to explain how moral concepts can flexibly perform these core functions in a continuously changing world.

One approach to meeting this challenge is to provide a psychological theory of concepts that can accommodate moral concepts and explain their role in motivating the selection of situated-action (i.e., the situationally appropriate and time sensitive selection of action) within dynamic

¹ I will adopt the traditional convention of denoting concepts in small capital letters, words in single quotes, and properties in italics (Laurence & Margolis, 1999; Machery, 2009; Bloch-Mullins, 2015).

moral experience.² Invariantism and Contextualism represents two groups of psychological theories of concepts (Löhr, 2017) that can be used to identify potential candidates. Traditional psychological theories of concepts are united in their commitment to an invariant representational format of concepts (Barsalou, 2012; Bloch-Mullins, 2015). According to traditional theories, concepts are invariant mental representations denoting some static ‘default’ or ‘core’ bundle of information (Bloch-Mullins, 2015; Machery, 2009). These concepts are stored in an individual’s long-term semantic memory and their stable cores are retrieved (or ‘tokened’) when facilitating high-level cognitive processes involving their referents (Machery, 2009). Traditional theories are contrasted with embodied (or ‘grounded’) psychological theories of concepts that reject the commitment to an invariant representational format. On embodied views, tokening a concept does not require retrieving a core collection of information stored in long-term memory independent from contextual features in the environment (Casasanto & Lupyan, 2015; Löhr, 2017; Michel, 2020; Yee & Thompson-Schill, 2016). Instead, concepts are contextually constructed to meet the demands of an agent given the situation and cognitive tasks (Casasanto & Lupyan, 2015; Löhr, 2017; Michel, 2020; Yee & Thompson-Schill, 2016). Embodied theories entail radically different perspectives on the representational format, conceptualization process, and function of concepts in moral cognition.

In this paper, I argue traditional psychological theories of concepts that posit invariant representational formats cannot provide a satisfactory theory of moral concepts. Although

² The distinction between psychological and philosophical theories of concepts is most notably introduced by Edouard Machery (2009) in *Doing Without Concepts*. Psychological theories of concepts are largely concerned with specifying their representational format (i.e., structure or lack thereof), acquisition, development, instantiation, and functional role within relevant cognitive capacities (Machery, 2009). These differ from philosophical theories of concepts whose goal is to explain the epistemic and metaphysical roles concepts play within propositional attitudes, such as beliefs and desires (Machery, 2009). In this paper, I will be solely concerned with psychological theories of concepts.

traditional theories can explain how moral concepts interpret sensory information and contribute to the formation of inferences like moral judgments, invariant representational formats are unable to accommodate the evaluative component of moral concepts while satisfactorily motivating situated-action within instances of moral experience. Instead, a flexible representational format is better suited to meet the unique context-specific features that arise during these instances. I will highlight a promising alternative candidate, most notably put forth by Casasanto and Lupyan (2015), that understands instances of moral concepts as “ad hoc” mental representations that are dynamically constructed to meet the needs of the specific situational context and the agent’s cognitive tasks.

Section 1 motivates the need to develop an empirically plausible psychological theory of moral concepts and clarifies both the descriptive and evaluative components of moral concepts. In Section 2, I will further clarify the Invariantism-Contextualism dichotomy with an emphasis on traditional psychological theories and propose how they can be plausibly extended to include moral concepts. Section 3 presents a challenge to traditional psychological theories of concepts by introducing an instance of moral experience to illustrate that the evaluative component of moral concepts cannot be invariant. I will argue a viable theory of moral concepts must allow for a flexible representational format and highlight Casasanto and Lupyan’s (2015) ad hoc conceptual theory as a plausible alternative in Section 4. Finally, I will conclude by considering this argument’s implications for research in moral cognitive science.

Section 1: The Features of Moral Concepts

In the last 20 years, moral cognitive science has experienced a renaissance with substantial progress made towards identifying and understanding the various psychological, cognitive, and neuroscientific processes that give rise to our moral thought and behavior (Cushman et al., 2017;

Greene, 2015; Heyes, 2021). Early research primarily focused on moral judgment formation and decision-making (Greene et al., 2001; Haidt, 2001; Sinnott-Armstrong & Wheatley, 2013), but as the field grew, research programs branched out to explore other topics and phenomena such as the acquisition of moral values and rules (Cushman et al., 2017; Nichols, 2021), the roles of emotion and affect in moral cognitive processes (Huebner et al., 2009; Prinz, 2006), the cognitive development (Heyes, 2021; Turiel, 2015) and evolutionary origins for our moral cognitive capacities (Heyes, 2021; Tomasello, 2016; Stanford, 2018), and the implications of moral content on our perceptual faculties (Gantman & Van Bavel, 2015; Gray et al., 2012; Sevinc et al., 2017). Alongside this growth, a variety of different, and sometimes incompatible, explanations for the phenomena in moral cognitive science emerged. There is now a growing need to begin bridging the high-level computational and behavioral explanations found in psychology with the low-level explanations provided by research in cognitive neuroscience. Computational modeling has recently been proposed as a promising methodological strategy to connect explanations across these levels of analysis (Crockett, 2016). However, if moral cognitive science aims to provide mechanistic explanations for the various multilevel processes underlying our moral cognitive capacities, it is crucial that the field *also* examines the kinds of representations that can plausibly instantiate these high-level computational and psychological processes into the human neurocognitive architecture.³ Since concepts are those mental representations that facilitate high-level cognition, generating multilevel neurocognitive explanations requires appealing to empirically plausible concepts. Therefore, the development of a satisfactory psychological theory of moral concepts that clarifies the representational features

³ Focusing on the neurocognitive representations is necessary for the field to begin adapting to what Boone and Piccinini (2016) call “the cognitive *neuroscience* revolution” (p. 1513). As pointed out by Boone and Piccinini (2016), explanations of multilevel neurocognitive mechanisms that comprise our cognitive capacities – such as those underlying moral cognition – are in terms of representation and computation.

of moral concepts is crucial to advancing the neurocognitive scientific study of morality. The goal of this section is to examine the representational features of moral concepts which enable them to represent morally relevant information and facilitate relevant processes. Through clarifying these features, potential constraints can be identified to determine and assess viable candidate psychological theories of concepts that may be considered when developing a theory of moral concepts.

Section 1.1: The Descriptive and Evaluative Components of Moral Concepts

The formation of a moral judgment, expression of a moral sentiment, performance of a moral action, and recognition of moral harm all require moral concepts to represent information containing morally relevant content – such as the predicted intentions of an agent, the actual outcomes of an action, the current structure of a social institution, or the possible implications of applying a specific moral norm to a given event.⁴ But what *kind* of information do moral concepts represent? In the case of the object concept CHAIR, it is clear the kind of information represented is primarily visual or information exogenous to us (exteroceptive). This information may also be multisensory – we might experience some affect when nostalgically reminiscing about our child’s high-chair by attributing to that collection of information some “affective significance” (Pessoa, 2009, p. 160).⁵ When we form a moral judgment containing CRUELTY, the primary kind of information represented is likely affective or endogenous information in us (interoceptive). However, in instances of tokening CRUELTY, the represented information is

⁴ For our purposes, the moral relevance of some set of information is relative to a specific cognitive agent and their situation, cognitive tasks, past experiences, etc. For example, if someone does not attribute moral standing to non-human animals, then they may not represent any moral concepts when slaughtering a pig. However, a bystander or a philosopher may still represent moral concepts when witnessing their actions or contemplating the event.

⁵ In moral philosophy and moral psychology, the terms ‘affective information’ and ‘emotion’ are often used interchangeably. However, in this paper, ‘affect’ will refer to interoceptive information with positive or negative value to the agent (Pessoa, 2009). Also, I will take ‘emotions’ to be high-level mental states and concepts that are tokened by an agent (Barrett, 2017; Cameron et al., 2015).

likely multisensory – there might be some combination of information about the observed actions of an agent alongside our feelings about these actions. The ability to represent both exteroceptive and interoceptive elements of experience are distinctive features of moral concepts.

A core feature of moral concepts is their ability to evaluate information. For example, imagine standing in line at the coffee shop and witnessing someone skip to the front immediately upon entering the building. In this experience, numerous concepts are tokened to represent salient sensory information. Some of these concepts are used to meaningfully interpret information about objects or features in the environment, but others may likely contain evaluative moral content. There will likely be an affective experience accompanying your observation of this individual's actions. This affective experience appears directly connected with the negative evaluation of their actions and the tokenization of a moral concept that roughly corresponds to WRONGNESS or BADNESS.⁶ Additionally, when communicating this experience to others, you might use terms like 'wrong' or 'unfair' to convey your evaluation of the individual's actions or the resulting state of affairs. This evaluative component is thought to be a characteristic feature of moral concepts (Fingerhut & Prinz, 2018; Väyrynen, 2021). By tokening a moral concept, a judgment is imposed onto some information or state of affairs in the world that extends beyond the current exteroceptive information (Fingerhut & Prinz, 2018; Väyrynen, 2021).⁷ In tokening WRONGNESS to evaluate the individual's actions, the attribution

⁶ The act of witnessing a moral violation does not always result in the conscious representation of a linguistic label such as 'wrong'. Thus, in this scenario, it is difficult to identify the mental representations resulting from the moral judgment with a precise linguistic label or word. This is primarily because concepts are not words (Casasanto & Lupyan, 2015). Concepts are mental states, whereas words are linguistic products used in communication. Despite sharing a close relationship, they are not identical (Casasanto & Lupyan, 2015).

⁷ In philosophy, this evaluative component is considered a feature of the information conveyed when using a moral term or moral concept, it is not thought to be a mental process of evaluation (Väyrynen, 2011). This is because concepts are being understood philosophically (see Footnote 2). Psychological theories will attempt to connect this evaluative component in the affective information contained within a token instance of a moral concept.

of moral significance is not a perceived property of any object or feature in the experience (Fingerhut & Prinz, 2018). Instead, it is the representation of salient interoceptive information which directs our attention, modulates executive functioning, and motivates goal-directed processes (Pessoa, 2009).

In addition to containing evaluative content, some moral concepts (if not all) are thought to possess an additional descriptive component (Väyryen, 2021). Moral concepts, such as JUSTICE, COMPASSION, or CRUELTY, possess both descriptive and evaluative information. These moral concepts function to categorize information and features in the experience while also attributing to that information some degree of moral significance. Consider a bystander witnessing you return the customer's dropped wallet in the initial scenario. This bystander may either represent you or your virtuous actions by tokening a moral concept like KINDNESS. Here, there is an evaluation of some state of affairs, but tokening KINDNESS also represents information beyond the evaluative content which can be used to construct additional inferences or moral judgments. In this scenario, tokening KINDNESS carries information about your character which the bystander may use to update their impressions of you or strengthen their expectations of appropriate action in similar situations.

The recognition of these two components in contemporary moral philosophy has given rise to an influential distinction between 'thick' and 'thin' moral concepts (Väyryen, 2021). Thin moral concepts (e.g., APPROPRIATENESS, RIGHTNESS, or BADNESS) are characterized as possessing only an evaluative component, whereas thick moral concepts possess both (e.g., KINDNESS, FAIRNESS, and CRUELTY) (Väyryen, 2021). In this paper, the thick-thin

distinction is understood as one of degree, rather than defining two separate natural kinds.⁸ For example, some moral concepts may possess a high-degree of descriptive content with low-degrees of evaluative content, others may contain low-degrees of descriptive content with high-degrees of evaluative content. I will use ‘thin’ to refer to those moral concepts possessing low-degrees of descriptive content (or none at all) and ‘thick’ will refer to all other moral concepts.

Section 1.2: Moral Concepts and Moral Motivation

A core feature of our moral judgments is their ability to motivate actions and conform our dispositions in accordance to acquired moral rules, norms, and intuitions (Nichols, 2021). A moral judgment is commonly understood as a mental state corresponding to an endorsement of a proposition containing moral content (Sinnott-Armstrong & Wheatley, 2013). Moral concepts can both enter into the processes that form moral judgments and result from the formation of a moral judgment, such as in the instance of tokening WRONGNESS. For the purposes of this paper, moral judgments do not need to be explicit in conscious reasoning and do not need to be clearly distinguished from other inferential processes containing moral content. Although both the descriptive and evaluative components of the moral concepts contribute to the formation process, the evaluative components of the moral concept play the primary role in moral motivation and the subsequent actions or dispositions resulting from the formation of moral judgments. For example, the moral judgment that motivates us to return the wallet appears to be influenced by the evaluative force contained within concepts representing the customer’s wellbeing.

⁸ Whether there are purely thin moral concepts is a matter of significant philosophical debate (Väyryen, 2021), but one’s stance on this issue is irrelevant to developing this paper’s main argument.

However, the actions we select across each token instance of WRONGESS are not formulaic. Instead, the motivational force and subsequent actions selected appear highly contingent on the situation itself, such as the constraints the situation places on our actions, the norms of appropriate behavior, and the goals of the agent. This gives rise to the following questions: Does each token instance of a moral concept possess the same degree of descriptive and evaluative content? What determines the magnitude of a moral concept's evaluative content during a specific token instance? Is this magnitude consistent across each token instance or do different contexts produce variation? How do these components combine in unique moral experiences to produce situated-action? These questions concern the content of a moral concept and its representational format, namely whether the content of a moral concept is fixed or sensitive to the situation's context. Determining this feature of moral concepts requires either developing a psychological theory of concepts or presupposing commitment to an already existing account. This section identified two representational features of moral concepts that can be used to assess the viability of candidate theories. Since these features concern the kind of information moral concepts represent, determining the set of viable theories requires examining their posited representational formats.

Section 2: The Invariantism-Contextualism Dichotomy as Candidates

Throughout modern cognitive science, numerous psychological theories of concepts have been proposed to explain the acquisition, storage, development, and deployment of concepts during both high-level and low-level cognition (Barsalou, 2012; Bloch-Mullins, 2015; Laurence & Margolis, 1999; Machery, 2009). To assess the viability of a given psychological theory of concepts, the central feature to consider is the representational format of concepts they posit. Three dichotomous properties are used to characterize the representational format of concepts:

amodal/modal, symbolic/non-symbolic, and static/situated (Barsalou, 2012). Concepts are amodal if their storage or representation are carried out in a conceptual system distinct from the sensory modalities of the assumed neurocognitive architecture (Barsalou, 2012; Machery, 2009). These are contrasted with modal concepts that are instantiated within the sensory modalities themselves (Barsalou, 2012; Machery, 2009). The symbolic status of concepts also shares a close connection to the assumed neurocognitive architecture. Symbolic concepts are structured, abstract mental representation that interpret and stand-in for sensory information that is recorded, or processed, within our sensory modalities (Barsalou, 2012). Non-symbolic concepts are non-structured mental representations grounded within the modality-specific systems or distributed throughout the neurocognitive architecture that instantiates the conceptual system (Barsalou, 2012). Finally, concepts are static if the mental representations in the neurocognitive architecture are either stored or retrieved independently of the situational context (Barsalou, 2012; Bloch-Mullins, 2015; Casasanto & Lupyan, 2015; Machery, 2009; Michel, 2020). Situated concepts differ from their static counterparts by taking a concept's content to contain contextual information specific to the given situation (Barsalou, 2012; Bloch-Mullins, 2015; Casasanto & Lupyan, 2015; Machery, 2009; Michel, 2020; Yee & Thompson-Schill, 2016). Despite differences among the various conceptual theories that have traditionally dominated modern cognitive science, most traditional theories take concepts to be static or invariant (Barsalou, 2012; Bloch-Mullins, 2015; Casasanto & Lupyan, 2015; Machery, 2009; Michel, 2020). Thus, this last dichotomous property is of significant interest to developing a psychological theory of moral concepts and can be used to partition existing theories into two categories: Invariantist and Contextualist (Löhr, 2017). This section aims to clarify the Invariantism-Contextualism debate and introduce traditional psychological theories of concepts while highlighting their common

commitment to an invariant representational format. Additionally, this section considers one plausible attempt to incorporate moral concepts into traditional theories.

Section 2.1: Clarifying the Invariantism-Contextualism Debate

According to Invariantist psychological theories of concepts, the capacity to meaningfully interpret and represent sensory information is a consequence of a concept's content being a stable core collection of information (Bloch-Mullins, 2015; Löhr, 2017). The ability to represent a wallet in our perceptual experience is done by applying the previously acquired concept WALLET, whose content is some set of characteristic features applying to all, or most, instances of the object. Invariantists argue this not only helps us to pull from our past experiences and learned associations with similar objects and similar situations, but conceptual stability enables the extraction of important details from a barrage of information that drastically varies in each experiential instance. An invariant concept's content is a consistent *default* collection of information that is present across each token instance of the conceptual category (Machery, 2009). This conceptual stability aids us in navigating throughout our dynamically changing environments (Bloch-Mullins, 2015; Löhr, 2017).

Previous attempts to define Invariantism (see Barsalou, 2012; Bloch-Mullins, 2015; Casasanto & Lupyan, 2015; Löhr, 2017; Machery, 2009; Michel, 2020) identify either the agent's stored concepts or the concept's retrieval process as being invariant. Although these previous definitions are sufficient to evaluate psychological theories of concepts positing a modular central processing unit as Invariantist, additional clarification and details are needed to determine whether theories positing either alternative architectures, such as connectionism, or pluralistic representational formats are also Invariantists. For example, on a connectionist framework, a concept represents information that is distributed throughout the network's nodes

(Barsalou, 2012) making it difficult to interpret ‘storage’ or ‘retrieval’. Additionally, psychological theories of concepts that posit a plurality of representational formats (Weiskopf, 2009) may or may not be considered Invariantist, since different mental representations may be stored or retrieved in different manners. To provide additional clarity and details, this paper will adopt the following definition:

Invariantism: A psychological theory of concepts is *invariant*, and posits an invariant representational format, if the conceptualization process (e.g., representation, tokenization, retrieval, construction, etc.) of a concept is performed by either accessing some stable collection of information from memory stores (e.g., retrieving a conceptual core) or activating some consistent set of modality-specific systems independent of the agent’s situational context (e.g., environmental constraints, the agent’s tasks or goals, or the agent’s expectations, etc.).⁹

This definition can be applied to both pluralist psychological theories of concepts and most plausible neurocognitive architectures. A pluralistic psychological theory of concepts is invariant if there is a type of concept tokened in this manner. Additionally, connectionist accounts can be Invariantist if the activation of nodes in the network are not sensitive to the agent’s situational context.¹⁰ Although some definitions of Invariantism give rise to a graded distinction (Michel, 2020), this definition creates a strict partition that is helpful for assessing candidate theories. If the conceptualization process is sensitive to the agent’s situational context, then that concept is

⁹ Michel defines ‘context’ in the following manner: “as a set of parameters that describe the environment in which a cognitive agent is embedded and which influence the exercise of cognitive competencies” (p. 625-626).

¹⁰ You might doubt whether any connectionist psychological theory of concepts is invariant in this manner. Regardless, I want to leave the possibility of Invariantist connectionist or embodied theories open.

not invariant. As will be shown in Section 2.2, this definition is not overly restrictive and accommodates theories traditionally thought to be Invariantist.

Contextualist psychological theories of concepts reject the claim that a concept's content or the conceptualization processes is independent of the agent's situational context (Barsalou, 2012; Bloch-Mullins, 2015; Casasanto & Lupyan, 2015; Löhr, 2017; Michel, 2020; Yee & Thompson-Schill, 2016). As Yee & Thompson-Schill (2016) state, "*the concepts themselves are inextricably linked to the contexts in which they appear*, so much so that the dividing line between concept and context may be impossible to clearly make out" (p. 1016). Contextualist theories reject the notion of a conceptual core, instead, each token instance of a concept is uniquely constructed to meet the demands of the agent's situational context (Casasanto & Lupyan, 2015; Michel, 2020; Yee & Thompson-Schill, 2016). The interpretation of a particular set of sensory information corresponding to a particular wallet tokens an instance of WALLET, whose content contains relevant details determined by perceived situational cues (Barsalou, 2009; Barsalou, 2012; Casasanto & Lupyan, 2015; Yee & Thompson-Schill, 2016). The conceptualization process combines previously acquired information of similar objects, learned associative information that is expected to be relevant given our past experiences, and relevant sensory information about the current situation (Barsalou, 2009; Casasanto & Lupyan, 2015). The resulting concept is a unique mental representation that is situationally constructed for the agent to meet their current tasks (Barsalou, 2009; Barsalou, 2012). Although there might be highly correlated sets of information across instances of experience, the information being represented is never the same collection of information and statistical correlations are not sufficient reasons to posit invariant cores to the concepts tokened (Casasanto & Lupyan, 2015). To clarify the proposed Invariantism-Contextualism dichotomy, I define Contextualism as follows:

Contextualism: A psychological theory of concepts is *contextual*, or *situated*, and posits a context-dependent representational format, if the selection of stored information or activation of modality-specific systems during the conceptualization process is sensitive to the agent's situational context, or the content of the token instance contains information of the current situational context.¹¹

Any psychological theory of concepts not in the extension of Invariantism will be considered Contextualist theories. A given account can be classified into either of these categories independently from other commitments concerning the representational format of concepts and neurocognitive architecture that instantiates the conceptual system.

Section 2.2: Traditional Psychological Theories of Concepts

Most traditional psychological theories of concepts originate from the “Good-Old Fashioned Artificial Intelligence” (GOF AI) approach in cognitive science (Barsalou, 2012). Although a variety of psychological theories of concepts can be classified under this umbrella, three theories are normally thought to comprise the traditional approach: prototype (or semantic memory), exemplar, and Theory-Theory (Bloch-Mullins, 2015; Machery, 2009; Michel, 2020).¹²

According to prototype theory, a conceptual category corresponds to a prototype representation containing a distilled collection of information of common properties and features extracted from category exemplars (Barsalou, 2012; Bloch-Mullins, 2015; Laurence & Margolis, 1999). To

¹¹ It is important to emphasize situational features contained in the content of a concept need to be about the *current* situational context, rather than previous contexts. As will be discussed in Section 2.2, exemplar theories posit exemplar memories that contain contextual details present during their acquisition. But exemplar theories still posit stable conceptual cores and are Invariantist (Barsalou, 2012).

¹² If using time periods, one can also lump connectionist theories of concepts into this category (see Barsalou, (2012) for a review of these approaches). However, connectionist theories differ substantially from the three theories listed in terms of their representational format, neurocognitive architectural commitments, and origination from the GOF AI approach. Additionally, since their initial introduction, the influence of connectionist theories varied and only recently ascended to dominance (Barsalou, 2012).

determine if an object, property, or feature is an instance of a particular concept category, a similarity calculation is performed by comparing that instance to the prototype (Bloch-Mullins, 2015). Exemplar theories identify a conceptual category with a set of stored exemplar memories that contains not only features common to category members, but also “idiosyncratic information about category instances along with details about the situations in which they occur” (Barsalou, 2012, p. 248). Like prototype theories, to determine membership in a conceptual category, a similarity calculation is performed between the target instance and a set of exemplar memories (Bloch-Mullins, 2015). Finally, Theory-Theory accounts take conceptual categories to represent theories of knowledge containing a “body of propositions that articulate people's knowledge within a given domain” (Laurence & Margolis, 1999, p. 44). To determine membership under Theory-Theory, a member would need to possess the same “causal structure” (Bloch-Mullins, 2015, p. 941) of other members within that theory (Laurence & Margolis, 1999).

Although prototype, exemplar, and Theory-Theory approaches are considered traditional psychological theories of concepts, substantial differences can be found among and within them. For example, prototype and Theory-Theory approaches usually require a modular neurocognitive architecture with a distinct conceptual processing unit which contain concepts and the processes that operate over them, but exemplar theories do not require a distinct conceptual processing unit and some versions take exemplars to be instantiated within the sensory modalities (Barsalou, 2012). Additionally, differences among the properties attributed to the representational format of concepts can also be found. Concepts under prototype and exemplar accounts are symbolic, but Theory-Theory approaches typically posit non-symbolic concepts (Carey, 2009; Laurence & Margolis, 1999). Finally, each theory is not homogenous, researchers working under one of these theories may adopt a variety of positions concerning the structure of concepts and the kind of

neurocognitive architecture that instantiates the human conceptual system (Laurence & Margolis, 1999).

Despite these differences, traditional psychological theories of concepts are united in their commitment to an invariant representational format (Bloch-Mullins, 2015; Michel, 2020). An important feature of traditional approaches is that concepts are *possessed* by the agent (Machery, 2009).¹³ After a concept's acquisition, it is stored within the agent's long-term semantic memory (Barsalou, 2012; Machery, 2009). Concepts still change and develop over time, but these modifications are incorporated into the stored mental representations diachronically (Machery, 2009). The conceptualization process tokening a concept consists in two stages: retrieval and application. When a cognitive capacity is being performed, such as the process of categorizing sensory information in perceptual experience, there is a retrieval of relevant concepts (e.g., prototype, set of exemplar memories, or cluster of propositional knowledge) from memory stores that are manipulated by relevant processes (Machery, 2009). Since no perceptual object, property, feature, or event is the same across each instance, the application process manipulates these representations to fit the situational context by recruiting relevant "background knowledge" (Machery, 2009, p. 12). A central task facing traditional psychological theories of concepts is to explain how these processes and mechanisms enable an invariant conceptual core to fit a specific situation through the incorporation of background knowledge (Machery, 2009). By appealing to background knowledge and relevant processes, traditional theories of concepts can (at least in

¹³ This notion of possession is highlighted by Casasanto and Lupyan (2015). On their ad-hoc approach, concepts are best understood as "something we *do with* the mind" (Casasanto & Lupyan, 2015, p. 546), not something that is possessed in the mind. This distinction is crucial as the possession of a mental representation likely drives us to think of conceptual categories being best represented by a stable core. Section 4 develops this point in more detail.

principle) meet the demands of a consistently changing environment and the agent's situational context.

To obtain a better understanding of traditional psychological theories of concepts and their ability to accommodate the dynamics of everyday experience, consider our initial coffee shop scenario and the recognition of the hurried customer's wallet. As we gaze around the room, our sensory faculties encode the incoming perceptual information. The interpretation of information is carried out by high-level cognitive capacities, such as the categorization of salient sensory information. To interpret the collection of sensory information corresponding to the perceived wallet, the conceptualization process begins by retrieving relevant mental representations stored in memory.¹⁴ The retrieved representations might be a prototype containing commonly occurring properties that have been acquired throughout various experiences with wallets, a set of exemplar memories of wallets, or a body of propositional knowledge about wallets and objects that fulfill similar functions. Regardless of the mental representation stored, they are retrieved independently of the situational context – the same prototype is retrieved in this instance and will be retrieved in a future instance unless changes are made between these periods, the same set of exemplar memories are retrieved from memory or the same modality-specific systems are reenacted to simulate these memories, or the same body of knowledge is retrieved from memory to identify the causal structure of the information that needs to be interpreted. During the application, various complex processes and mechanisms are carried out to incorporate background information, learned associations, and applying additional concepts to fit our invariant core to this particular situation. A full explanation of this application process is left to

¹⁴ How the agent and conceptualization process determine relevance is left to be explained by the specific psychological theory of concepts being considered. This is not an apparent problem for embodied theories as the cues that determine relevance are the same cues that construct concepts.

explained by developing a particular psychological theory of concepts and connecting it to research on perception, attention, cognitive control, etc. Whatever the explanation may be, these additional processes are relegated to background knowledge on traditional theories (Machery, 2009; Michel, 2020).

Although the above explanation is incomplete, it provides a rough outline of the various high-level processes involved in the conceptualization processes that tokens an object concept. Attempts to extend these traditional accounts to other cognitive domains requires explaining how the content of concepts can contain that kind of information and how those concepts function with the relevant processes facilitating those cognitive capacities. For example, in the case of auditory perception, an explanation would need to be given for how those concepts represented auditory information. Additionally, for perception of social phenomena, an explanation would need to explain how social information becomes represented. Therefore, if traditional psychological theories of concepts can accommodate moral concepts, the first step is to determine how moral content is contained within an invariant conceptual core and how these concepts function within processes underlying our moral cognitive capacities.

Section 2.3: Accommodating Moral Concepts in Traditional Psychological Theories of Concepts

Attempts to incorporate moral concepts into traditional psychological theories fall into two strategies: (1) *Containment* – attempts that explain how both descriptive and evaluative information are contained within a moral concept's content; or (2) *Background* – attempts that explain how a moral concept's content interacts with other neurocognitive systems instantiating these components. If traditional theories attempt to pursue a *Containment* strategy, then they must connect both the descriptive and evaluative content to empirical neurocognitive research

and explain how this information can be *contained* within the invariant content of a moral concept. If traditional theories attempt a *Background* strategy, then they must explain how the retrieved invariant conceptual core interacts with *background* processes and representations to describe and evaluate information. The *Containment* thesis maintains a direct connection between tokening a moral concept and motivation. The *Background* thesis explains this connection through appealing to background knowledge and relevant processes.

Section 2.3.1: Incorporating the Descriptive Component of Moral Concepts

Regardless of the strategy taken, the descriptive component of a moral concept is likely analogous to the descriptive content of concepts within other domains. When interpreting multimodal sensory information corresponding to an object concept, the conceptualization process will retrieve and apply acquired concepts from memory stores to that collection of information. Once a concept is selected, associative information and other information within the content of the concept is made available for use by other high-level processes. For example, tokening a chair carries descriptive content that might facilitate the selection of actions and the deployment of motor commands to accomplish action goals, such as those associated with sitting or walking towards the object. An analogous process is likely true for those moral concepts containing a descriptive component as well. When interpreting multimodal sensory information corresponding to a person, the conceptualization process may token the concept PERSONHOOD and the descriptive content of this concept might also give rise to the selection of action, the formation of inferences, or the tokenization of other composite concepts that contain the initial concept (e.g., BROTHER). Since the descriptive component of a moral concept functions similarly to the descriptive component of concepts in other domains, traditional theories of concepts can simply refer to explanations provided for those other concepts to explain how the

descriptive component of moral concepts are contained within an invariant conceptual core.

Thus, the challenge for traditional theories will be to accommodate the evaluative component of moral concepts.

Section 2.3.2: Incorporating the Evaluative Component of Moral Concepts

As discussed in Section 1.1, the evaluative component can be used to confer motivational features to moral concepts, enabling them to perform their core cognitive functions – such as directing our attention to salient moral content within our experience, forming moral judgments to help us navigate our social environments, and performing moral actions. The evaluative and motivational features of moral concepts can be understood by identifying this evaluative component with affective and interoceptive information. If traditional psychological theories of concepts pursue the *Containment* strategy, then they must explain how this information is contained within the content of a moral concept and enables the attribution of moral significance. If traditional theories pursue the alternative *Background* strategy, then an explanation is required for how moral concepts interact with other neurocognitive systems to perform their core functions.

The development of either strategy requires understanding how emotional and affective states relate to moral concepts and enable them to perform their core functions. During experience, salient sensory information directs attentional resources (Pessoa, 2009). Affectively significant stimuli (e.g., objects, features, elements, etc.) modulate executive control – dedicating attentional resources and prioritizing that stimuli in processing, sensitive to its valence, arousal, and perceived importance to the agent's tasks and goals (Pessoa, 2009). The affective value attributed to a given collection of sensory information is acquired through experience but is also sensitive to the agent's situational context (Pessoa, 2009). As cited in Pessoa (2009), various

brain regions are recruited to processing these stimuli – such as areas in the anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), and regions in the prefrontal cortex (PFC).

Additionally, Pessoa (2009) highlights that activation of the ACC sends signals and recruits other regions used for “[modulating] cognitive, motor and visceral responses” (p. 164). The processing of affective information is highly connected to motivation as well (Pessoa, 2009). For example, stimuli associated with reward also direct attentional resources and executive functioning to obtain or “maximize potential reward” (Pessoa, 2009, p.164). In this sense, emotional and affective states can be connected to motivational processing (Pessoa, 2009).

The processing of these stimuli likely contributes to the formation of concepts, which are also thought to direct the flow of attentional resources (Barsalou, 2009; 2012). The attribution of affective significance to stimuli is not unique to moral content, so the processing of affective information does not necessarily contribute to the formation of moral concepts. However, studies have shown that perceived moral content engages the ‘salience network’, directs attentional resources, and modulates executive control (Sevinc et al., 2017; Sevinc & Spreng, 2014). This suggests that affective information either contributes or constitutes moral concepts. Therefore, if the evaluative component of a moral concept’s content is connected to these emotional and affective states, then this component can be used to connect moral concepts to motivation and the core functions they perform in our moral cognitive capacities.

In philosophy and moral cognitive science, emotion and affect are thought to share a close connection with our moral cognitive capacities (see Fingerhut & Prinz, 2018; Haidt, 2001; Huebner et al., 2009; Kelly, 2020; Nichols, 2021; Prinz, 2006). For example, moral sentimentalism – one of the dominant positions in moral philosophy and psychology – argues moral judgments are shaped by our emotional and affective states (Kelly, 2020; Nichols, 2021;

Prinz, 2006). Some sentimentalists even take moral judgments (see Kelly, 2020) and moral concepts to be constituted by them (see Huebner et al., 2009).¹⁵ Other accounts, such as moral rationalism (May, 2019) or constructivism (Cameron et al., 2015), also take emotional and affective states to be tightly connected to moral cognition, but these alternative accounts differ regarding the contribution these states make to the tokenization of moral concepts and formation of moral judgments. But, as noted by Huebner et al. (2009) regarding moral judgments, these positions must specify the stage at which emotional concepts and affective states enter into the conceptualization process and the formation of moral judgments. Focusing on the conceptualization process, if emotional concepts and affective states constitute the tokenization of a moral concept, then the tokened moral concepts are simply the emotions and affective states. If emotional concepts and affective states are integrated within the conceptualization process, then these states are still contained within the content of moral concepts but might be combined with additional information. Finally, if emotional concepts and affective states result from tokening a moral concept, then the former are not contained within the latter and, instead, will result from interactions between moral concepts and other neurocognitive systems. To determine which of these conditionals best explains the relationship between moral concepts and emotional and affective states is beyond the scope of this paper, but one's position on these issues will impact the strategy taken to explain the evaluative component of moral concepts. If the first two conditionals are favored, then a traditional psychological theory of concepts must pursue the *Containment* strategy. If the third conditional is favored, then the *Background* strategy is

¹⁵ As stated in Section 2.3, the formation of a moral judgment can be understood as endorsing a proposition containing moral content. Thus, a moral judgment will token both moral and non-moral concepts. In some cases, they token characteristically thin moral concepts. Since the present task is to connect evaluation to emotion and affect, outlining these positions are important to consider for this paper's main argument and I will take the conceptualization process and the processes of forming moral judgments to be similar in this section.

sufficient. If one rejects the sentimentalist tradition, any strategy can be pursued given compatibility with the rest of that position's commitments.

Section 2.3.3: Outlining the Containment and Background Strategy

If traditional psychological theories of concepts pursue the *Containment* strategy, then both the descriptive (if any) information and evaluative information must be contained in moral concepts. During instances of experience and thought, the interpretation of morally relevant information will be identified, classified, and evaluated through the application of stored concepts. The conceptualization process retrieves relevant concepts from memory stores and applies them to the given situation. Since the content of a moral concept is invariant, the descriptive content of the tokened moral concept is also invariant. If necessary, background knowledge and relevant processes will be recruited to fit the invariant conceptual core to the agent's present situational context. Likewise, the evaluative content of the tokened moral concept is also contained within the invariant conceptual core and applied similarly. The *Containment* strategy fits well with the sentimentalist claim that emotional and affective states either constitute or primarily contribute to the tokenization of moral concepts. Since concepts contain an invariant representational format, these states and tokened moral concepts, or formed moral judgments, will share one-to-one correspondences (Kelly, 2020). In this sense, the representation of moral content will be causally linked to a specific emotional (or affective) state (Cameron et al., 2015).

Traditional psychological theories of moral concepts adopting the *Containment* strategy may appeal to either the highly influential *Basic Emotion* framework (see Cameron et al., 2015; Kelly, 2020) or the equally influential *Moral Foundations* theory (see Graham et al., 2013) to justify this isomorphic relation between moral concepts and discrete emotional and affective

states. On the Basic Emotion framework, each emotional experience is understood as a token instance of a specific universally shared and evolutionarily innate emotional category (Cameron et al., 2015). Distinct emotional categories evolved as specific behavioral response mechanisms which preselect specific actions and judgments enabling us to adaptively navigate external environments (Cameron et al., 2015). These evolutionarily innate mechanisms produce ‘habitual’ or ‘automatic’ responses that are consistent across each instance of emotion (Cameron et al., 2015). For example, the perception of a possible threat may token an emotional instance of ‘fear.’ This emotional reaction produces a habitual response by selecting a predetermined set of motor commands. This habitual response might produce observable physiological reactions, such as tensing of the facial muscles or directing blood flow to the agent’s extremities. Moral Foundations theory, like the Basic Emotion framework, takes instances of distinct emotional categories to produce specific habitual responses, however, it further argues these response mechanisms share an isomorphic relation to the formation of moral judgments about discrete moral categories (Graham et al., 2013). On this view, discrete moral categories comprise the ‘foundations’ of our moral cognitive capacities. Although these moral foundations are assumed to be innately acquired, the ontogeny and experience of the agent may influence the development of these categories and their associated responses (Graham et al., 2013). Since specific correspondences connect these emotional response mechanisms to distinct foundations that categorize instances of moral content, the formation of a moral judgment is ‘intuitive’ or ‘habitual’ (Graham et al., 2013).

An important consequence of these approaches is that each token instance of a moral conceptual category – e.g., WRONGNESS, KINDNESS, CRUELTY, etc. – will contain the same evaluative content and this invariant content will motivate the same response mechanisms,

which select some set of predetermined actions or dispose us to the represented information in the same manner. If the motor commands that instantiate the selected action plans or alter our disposition towards the represented information need alteration to meet the agent's situational context, then executive control will direct relevant processes to incorporate relevant background knowledge to alter these responses following the conceptualization process that tokens the moral concept. Through connecting the evaluative component of a moral concept directly to motivation, this strategy is well suited to explain how the formation of moral concepts disposes us to act and form impression about the represented information. However, due to the invariant representational format of traditional psychological theories of concepts, this strategy heavily relies on either the incorporation of background knowledge or habitual reactions – either as moral intuitions or as emotional responses – to explain our moral cognitive capacities.

The *Background* strategy does not require emotional and affective states to be contained within the content of a moral concept. Instead, these states are understood to be the result of other processes or concepts independent of moral concepts. On this view, a moral concept is still tokened to interpret morally relevant sensory information and the conceptualization process will still recruit an invariant conceptual core to classify this information. The descriptive and evaluative components of a moral concept are still contained within this invariant conceptual core, but the content of the moral concept is not used to classify or interpret present emotional or affective information. After the conceptualization process, the content of a moral concept may still produce emotional and affective states through interacting with other neurocognitive systems. This interaction post conceptualization can be considered an instance of fitting the moral concept to the situation via recruiting relevant background knowledge and processes. Moral motivation may still result from the tokenization of a moral concept, but understanding

this motivational force relies heavily on explaining the interaction of moral concepts with other neurocognitive systems. Although the *Background* strategy easily maintains the invariant representational format that is characteristic of traditional psychological theories of concepts, connecting moral concepts to their core functions – namely their motivational roles – represents a substantive challenge. As a result, many proponents of traditional psychological theories of concepts may wish to prioritize developing the *Containment* strategy over the *Background* strategy.

Traditional psychological theories of moral concepts appear well suited to explain many core features of moral concepts that enable them to facilitate high-level processes underlying our moral cognitive capacities. It is clear that traditional theories can, in principle, explain the interpretation of morally relevant sensory information and their roles in forming propositional attitudes within moral thought and reasoning. However, the invariant representational format entails a brittle conceptual core that must recruit background knowledge and relevant processes to meet the agent's situational context. The next section will examine the ability of traditional psychological theories of moral concepts to accommodate another central feature of moral concepts – the motivation of situated-action during dynamic moral experience.

Section 3: The Challenge of Moral Experience and Situated-Action for Traditional Psychological Theories of Moral Concepts

In this section, I will present a significant challenge for traditional psychological theories of concepts that attempt to provide a viable theory of moral concepts. The challenge concerns an instance of moral experience and the function moral concepts serve in motivating situated-action. Here, I argue this challenge results from traditional theories' adherence to an invariant representational format and these attempts fail in two respects: (1) they are unable to explain

variable moral motivation and evaluation that results from tokening moral concepts; and, (2) even if adjustments are made to avoid these issues, these theories are unable to preserve the motivational features of moral concepts through appealing to interactions between moral concepts and other neurocognitive systems.

Section 3.1: Introducing a Moral Vignette

Imagine walking down a familiar street late at night.¹⁶ Having walked this path numerous times before, traveling to your destination is an almost habitual task. You know what turn to make, when to move to the other side of the street, and you are aware of the likelihood that certain situations might be encountered (e.g., the possibility of being approached by a stranger, of seeing a parked car, etc.). So far on your stroll, nothing is out of the ordinary and your mind freely wanders, sometimes you are reflecting on tomorrow's tasks and other times you are attending to specific stimuli that become salient, such as the uneven sidewalk or a misplaced construction cone. However, as you round the corner of a familiar building, this mundane experience quickly changes – you witness a group of young kids douse an alley cat with gasoline and light it on fire. You are startled by this sudden and unexpected state of affairs. You form a seemingly 'automatic' or 'intuitive' moral judgment to evaluate the event by tokening a moral concept that can be roughly characterized as WRONGNESS which condemns the kids' actions. Your affective experience is in flux and you feel a variety of emotional instances that, in retrospect, may be described as 'disgust', 'fear', or 'anger.' There is an immediate and undeniable motivational force to engage in an action – any action – to disperse the kids and attempt to save the cat. As you scan your visual field, you notice a

¹⁶ This is an adaptation of Gilbert Harman's (1977) famous moral vignette outlined in *The Nature of Morality: An Introduction to Ethics*.

small rock near your feet and, without any conscious deliberation, you pick up the rock to throw it near the kids.

Although abnormal, this moral vignette can be used to illuminate the close connection shared between tokened moral concepts, evaluative content, moral motivation, and situated-action. In this experience, moral concepts serve central roles in interpreting perceptual information, forming moral judgments, and motivating our actions in response to perceiving the children's actions as cruel. For example, some of the moral concepts that contribute to our evaluations are thin moral concepts like WRONGNESS and thick moral concepts like CRUELTY, HARM, or WELLBEING. Additionally, these tokened moral concepts causally contribute to the acquisition of various principles, rules, and values that can be used to justifying our formed moral judgments or post-hoc reasoning about the situation. Our initial evaluation appears to be a consequence of our social expectations being violated – such as the appropriateness of gratuitous cruelty – and these norms and expectations are acquired early in our moral enculturation and continue to develop throughout our lives. Upon turning the corner, attentional resources are dedicated towards salient sensory information containing moral content. Our conceptualization processes tokens various moral and nonmoral concepts to evaluate perceived stimuli, the children's action, and motivate situated-action in response to the perceived moral injustice. The allocation of cognitive resources and processes through executive control is responsive to the affect-laden information within the experience and the conceptualization process appears to be carried out in a task and goal-directed manner that is sensitive to our situational context.

Any viable candidate theory of moral concepts ought to accommodate the application of moral concepts within dynamic moral experiences. However, I argue that traditional psychological theories of concepts that posit an invariant representational format are unable to

accommodate moral concepts in these situations. To develop this challenge, the first step is to clarify specific instances of moral concepts and their contribution to motivating the formation of our moral judgments and selected actions. Although this experience requires the tokenization of numerous concepts and many are likely to contain morally relevant content, I will focus on the token instance of WRONGNESS.¹⁷

The tokenization of WRONGNESS – a characteristically thin moral concept – results from the formation of our moral judgment about the event and children’s actions. The quick and seemingly automatic formation of our moral judgment provides an evaluation that is causally connected to the emotional and affective states present within the experience. The ability to perform a quick evaluation by tokening WRONGNESS enables us to adaptively respond given the situational context. These situational constraints and the perception of potential harm motivates the quick formation of a moral judgment and selection of actions in response to our moral evaluation. Since processing is prioritized for high-arousal affective stimuli (Pessoa, 2009) and the situational context places time-constraints on these processing mechanisms, this token instance of WRONGNESS is characteristically thin and possess a low-degree of descriptive content. Here, the content of WRONGNESS primarily contains evaluative information, such as our interoceptive information resulting from the recognition of the kids’ actions as cruel. Although the content of this concept is consciously accessible, explicit representation of a

¹⁷ As mentioned in footnote 6, moral concepts are mental states and the attribution of a linguistic label or symbolic expression is difficult to provide. One can easily interchange the symbolic expression ‘WRONGNESS’ with ‘BADNESS’ and the same conclusions can be reached. Casasantos and Lupyan (2015) argue confusion arises when the linguistic label or symbolic expression denoting the tokened moral concept is interpreted as a precise characterization of the concept and this confusion contributes to the default assumption that concepts possess an invariant representational format.

linguistic label and reflective deliberation is not required to motivate action selection. Instead, our subsequent behavior appears motivated by tokening WRONGNESS.

Alternatively, thick moral concepts can also be considered. For example, one instance of a thick moral concept is the tokenization of CRUELTY or INJUSTICE that causally contributes to the formation of our moral judgment and the conceptualization process of WRONGNESS. The descriptive component of this tokened thick moral concept categorizes the children's acts as an instance of CRUELTY and the evaluative component motivates the formation of our moral judgment by negatively disposing us towards the represented information. In this moral vignette, tokening CRUELTY, or any other thick moral concept, does not directly motivate the selection of actions. Instead, these thick moral concepts dispose us towards their content and contribute to the formation of judgments, construction of inferences, or the tokenization of other concepts.

Section 3.2: Moral Concepts Possess Variable Evaluative Content and Motivational Force

The central challenge facing any traditional psychological theory of concepts attempting to accommodate moral concepts can be posed as follows: if moral concepts possess an invariant representational format and their tokenization is independent of the agent's situational context, how can they motivate the selection of situated-action or dispose us to form situationally appropriate and time-sensitive judgments or inferences? Here, I will focus on the token instance of WRONGNESS, rather than instances of thick moral concepts like CRUELTY, to develop this challenge. However, before developing this challenge, a preliminary objection needs to be addressed regarding the selection of WRONGNESS over other possible moral concepts.

An objection can be advanced against focusing on a moral concept that results from the complex inferential process underlying moral judgment formation. One could argue this

challenge is merely against static, or formulaic, accounts of moral judgment formation. Furthermore, it could be argued that attributing a singular concept of WRONGNESS to the outcome of this formed moral judgment is an oversimplification and a formed moral judgment tokens numerous moral and nonmoral concepts, not just WRONGNESS.

I offer three defenses against this preliminary objection. First, on traditional psychological theories of concepts, high-level processes token concepts. When considering our cognitive capacities for morality, the inferential process of moral judgment is one example of a high-level process. Recall the definition of Contextualism provided in Section 2.1 – if the conceptualization process is sensitive to the situational context or the content of the tokened concept contains information of the current situational context, then that tokened concept is contextual. Now, since the processes underlying the formation of a moral judgment can be understood as analogous to the conceptualization process tokening this instance of WRONGNESS, then successfully illustrating this process as contextual is sufficient to also show the resulting moral concept tokened by this process is contextual. Second, moral concepts do not need to be primary or simple, they can (and likely do) emerge from a complex conceptualization process that integrates other conceptual and nonconceptual information. The common tendency to treat concepts as single units that can be precisely characterized with a unique symbolic expression or word does not entail that actual psychological concepts share these properties or, during instances of experience, clear delineations between concepts can be found. By drawing upon Section 1.2, this formed moral judgment is a mental state that roughly corresponds to endorsing the proposition, “Their actions are wrong” or “This event is wrong.” Here, this token instance of WRONGNESS confers evaluative and motivational force to the formed moral judgment. So, in this instance, identifying this mental state as a token instance of WRONGNESS is justified.

Finally, others (see Abend, 2011; Fingerhut & Prinz, 2018) have previously argued the robust descriptive component of thick moral concepts appear highly context-dependent. Fingerhut and Prinz (2018) argue token instances of JUSTICE – a characteristically thick moral concept – appear highly variable and they write, “There are many different ways to be just: distributing goods equitably, following procedures of due process and rectifying discrimination. These are not discrete, touchable entities, and they do not look alike” (p. 1). Since these arguments have already been proposed and not all moral concepts contain robust descriptive components, the third counter objection is the following: considering characteristically thin moral concepts will yield a more substantive challenge for traditional psychological theories of concepts.

Now, having addressed the prior objection, I will resume developing this challenge. Consider the tokening of WRONGNESS in the above moral vignette. According to traditional theories, the conceptualization process retrieves an invariant core collection of information that identifies the conceptual category of WRONGNESS from memory stores to interpret sensory information. The primarily kind of information contained within the content of this moral concept is evaluative. As outlined in the previous section, traditional theories have two options to accommodate the evaluative features of moral concepts: either they pursue a *Containment* strategy and take this evaluative information to be contained within the content of the concept, or they attempt a *Background* strategy and take this evaluative information as emerging from interactions between the tokened moral concept and other neurocognitive processes.

Consider traditional psychological theories of concepts that posit the evaluative information is contained within the content of the tokened moral concept. On this view, emotional and affective states contribute to the evaluation of salient stimuli. Emotional and affective information direct the allocation of resources and processing, shape our dispositions towards

stimuli, and motivate various action plans that determine our behavioral responses. But our everyday experience seems to suggest that not all token instances of WRONGNESS motivate us to form same dispositions or perform the same responses. For example, our decision to scan our immediate visual field and identify an object to meet the demands of our present tasks will not likely be our response when we evaluate and identify the actions of the customer who immediately proceeds to the front of the line upon entering the coffee shop by tokening WRONGNESS. Other token instances of WRONGNESS may also yield in no direct behavioral response. These examples appear to suggest a one-to-many relation holds between token instances of a moral conceptual category and the responses or dispositions they motivate. A one-to-many relation suggests this evaluative information varies across each token instance and is situated to the agent's situational context. If moral concepts possess an invariant representational format, then this variable evaluative information cannot be contained within the moral concept's conceptual core. However, this claim conflicts with the *Containment* strategy. Therefore, traditional psychological theories of concepts pursuing the *Containment* strategy fail to provide a satisfactory explanation for motivated action and dispositions when the selection of action and our dispositions are the direct consequence of tokening a moral concept.

One possible solution for these approaches is to deny the one-to-many relation. In advancing this solution, these attempts may appeal to either the Basic Emotion framework or the Moral Foundations theory to justify the rejection of this isomorphic relation. This resolution maintains Invariantism by claiming each token instance of a moral concept corresponds to either a discrete emotional response (if appealing to the Basic Emotion framework) or an 'innate' or 'automatic' emotional or affective evaluation (if appealing to the Moral Foundation framework). Relevant

background knowledge and additional processes aid in the application of a moral concept and will tailor our subsequent behavior to the agent's situational contexts.

Unfortunately for traditional theories attempting to pursue the *Containment* strategy, this alternative explanation is also unsatisfactory. To understand how it fails, consider some of its consequences. On this view, the dynamic selection of action results from specific emotional responses like 'fear' preselecting action plans which activate specific motor commands. The preselection of action plans is necessary to maintain an invariant conceptual core and enable the agent to perform situated-action. If there is an isomorphic relation between tokened moral concepts and either discrete emotional responses or intuitive evaluations, then tokening two instances of the same moral conceptual category ought to motivate the same actions or dispositions. However, the actions selected in this moral vignette and other ecologically valid scenarios are not 'reflexes' and our behavioral responses are not formulaic. Instead, our responses appear 'intelligent' and appropriately selected to adaptively engage within our current environments. Additionally, it is unlikely the actions selected result from 'automatic', 'innate', or 'intuitive' emotional or affective evaluations.

The existence of these isomorphic relations becomes increasingly dubious when considering recent empirical evidence in motor control theory which suggests even low-level motor commands are highly sensitive to an agent's situational context and intentions (Barrett & Finlay, 2018; Fridland, 2017). For example, Barrett and Finlay (2018) survey empirical behavioral ecology literature on the survival behaviors of larval zebrafish to argue even commonly thought 'reflexive' behavioral responses are highly sensitive to the organism's situational context. Additionally, the complex action plans implementing these survival behaviors are best understood as instances of "purposeful motor actions" (Barrett & Finlay, 2018, p. 174).

Zebrafish are appropriate examples to illustrate this point as they are “approximately 5 mm long, virtually experience-free, and are heavily preyed upon” (Barrett and Finlay, 2018, p. 174). Additionally, they also possess a premotor cortex that integrates multimodal sensory information to issue motor commands implementing adaptive behavioral responses upon perception of a threatening stimuli. When situational features place constraints on the duration of time available for processing sensory information and selecting an action plan, zebrafish will quickly (within a matter of 24 milliseconds) change their movement to the opposite direction and travel away from the perceived threat (Barrett & Finlay, 2018). However, Barret and Finlay (2018) highlight these selected actions are not formulaic or rigid reflexes – when situational features of the environment allow a longer duration of time to plan a behavioral response, zebrafish will implement increasingly more complex and evasive action plans where even “places to hide and other environmental affordances can be integrated into the decision” (p. 174). This evidence suggests even low-level motor commands are sensitive to variable features in the agent’s situational context and can occur without possession of a complex conceptual system.

Now, consider conceptual agents where behavioral responses are casually connected to concepts. If tokening a concept directly motivates the selection of an action plan containing these context-dependent motor commands, then that concept cannot possess an invariant representational format. In the moral vignette introduced above, the action plan to search our visual field contained context-dependent motor commands that were sensitive to features of the current situational context and our intentions. If we consider alternative scenarios that token instances of WRONGNESS, these instances will also motivate the selection of actions or the formation of dispositions that are sensitive to the current environment and its unique situational features. Since motivation is largely connected to the emotional and affective evaluation of

sensory information, traditional theories cannot maintain the *Containment* thesis while also positing an invariant representational format. This implies traditional theories are forced to consider a *Background* strategy if they wish to accommodate instances of moral concepts that motivate situated-action.

According to traditional psychological theories of concepts that pursue a *Background* strategy, the evaluative content that motivates our actions and dispositions is not contained within the stored invariant core of a moral conceptual category. Instead, the evaluative content and motivation results from the interaction between moral concepts and other neurocognitive processes. However, if the evaluative features of a moral concept are not contained within the content of the conceptual category, then these views must identify a moral concept with their descriptive component and argue that the evaluative content is contained in other non-conceptual mental representations and processes. This approach allows moral concepts to remain invariant while still permitting variable evaluative content and motivational force across each token instance of a moral concept. Here, the conceptualization process will retrieve an invariant moral concept and its tokenization, alongside the tokenization of other non-moral concepts, will interact to produce non-conceptual emotional and affective processes that ultimately drive our actions and responses. However, if these approaches hope to remain invariant, any mental representations that enable these background affective processes must be non-conceptual or else they will contain context-dependent representational formats.

But these views carry an unsatisfactory consequence undermining their viability as potential candidate theories to provide a satisfactory psychological theory of moral concepts. If moral concepts no longer contribute an evaluative judgment upon their tokenization and represent evaluative content during experience, they can no longer directly motivate action selection or

shape our held dispositions about this content. Additionally, moral concepts that solely contain descriptive content are unable to accommodate to instances of thick moral concepts, like the token instance of CRUELTY which contributes to the formation of our moral judgment in the target moral vignette.¹⁸ Tokening CRUELTY interprets that bundle of information corresponding to the children's actions and motivates the formation of our moral judgment which evaluates the scenario and their actions as wrong. However, since tokening CRUELTY does not contain those variable emotional and affective states, these accounts still struggle to explain how CRUELTY contributes to the formation of our moral judgment. Instead, it appears that our moral judgment is merely a consequence of nonconceptual processing and almost entirely shaped by our emotional and affective states. Although this conclusion may not seem contradictory to mainstream sentimentalist traditions in moral psychology, it results in either the elimination of moral concepts or relegates them to an inconsequential status during moral experience. However, our formed moral judgments do appear shaped by background conceptual knowledge and acquired moral rules (Nichols, 2021).

Regardless of how the evaluative component of moral concepts is explained, traditional theories of concepts fail to serve as viable candidates to provide a satisfactory theory of moral concepts. This failure is a direct consequence of traditional theories positing an invariant representational format which takes the content of a moral conceptual category to be static across all token instances. Thus, a satisfactory theory of moral concepts ought to consider psychological theories of concepts that are Contextualist.

¹⁸ Since these approaches will need to deny purely thin moral concepts, they will also reject the claim that our formed moral judgment tokens an instance of WRONGNESS. Instead, the notable moral concepts in the scenario are those characteristically thick moral concepts and a viable theory of moral concepts ought to explain how they causally contribute to the moral judgment that is formed.

Section 4: Context-dependent Representational Format and Ad Hoc Concepts

As illustrated in the previous section, a satisfactory psychological theory of moral concepts must permit a flexible representational format to accommodate the variable evaluative content and motivational force contained within each tokened moral concept. Although this argument's conclusion encourages considering alternative psychological theories of concepts beyond Invariantism, it does not privilege one alternative theory over others. Questions still remain regarding the specific representational format of concepts, the neurocognitive architecture of our conceptual system, and the roles moral concepts play in capacities like attention, perception, memory, and more. However, the dominance of the traditional approach has made it difficult to consider alternative accounts that posit a radically different conceptual structure and do not fit the standard stimulus-response paradigm (Hutchinson & Barrett, 2019).¹⁹ But if a concept is not a static mental representation stored in long-term memory that becomes retrieved during the conceptualization processes to facilitate high-level cognitive capacities, then what is it? What properties enable flexible concepts to satisfactorily motivate situated-action within instances of moral experience? To answer these questions requires examining a specific Contextualist theory – not all Contextualist accounts characterize the conceptual system in the same manner, nor do they posit representational formats with the same properties. Contextualist theories could posit a modular cognitive architecture or posit concepts that are amodal or symbolic. However, most of the dominant Contextualist theories draw heavily upon embodied tradition, where concepts are instantiated within the modality specific systems themselves. To consider the implications of this

¹⁹ See Hutchinson and Barrett (2019) for a brief overview of the traditional stimulus-response paradigm and its traditional dominance in the cognitive sciences. Historically, traditional psychological theories of concepts have assumed this paradigm. For example, as the conceptualization process that tokens a concept is merely the interpretation of perceived sensory information, the tokened moral concepts entering into background processes, and other representations to produce an outcome, such as a behavioral response.

paper's main argument for research in moral cognitive science and to motivate the viability of developing a Contextualist theory of moral concepts, I will briefly consider one promising approach in this section – the ad hoc conceptual theory of Casasanto and Lupyan (2015).

Section 4.1: Introducing Ad Hoc Concepts

Casasanto and Lupyan (2015) argue for an ad hoc conceptual theory that radically differs from traditional psychological theories of concepts. On this view, concepts are unique mental representations constructed to meet the agent's current situational context. Rather than identifying concepts with stored mental representations that are acquired through experience, possessed by an agent, and retrieved to interpret sensory information, concepts are identified with the conceptualization process itself. These constructed mental representations are not stored as such in the agent's memory or possessed by the agent in a specific conceptual processing unit, instead, as Casasanto and Luypan (2015) write, "they are something we *do with* the mind" (p. 546). The conceptualization process tokening an ad hoc concept recruits relevant non-conceptual and associative information that the agent has acquired through past experience and constructs a mental representation enabling the agent to actively engage within their current and anticipated situational context. The relevance of this stored information for the agent's current situational context is determined by "internally generated or external cues" (Casasanto & Lupyan, 2015, p. 546) that can either be identified consciously or unconsciously among the currently perceived or anticipated sensory information. Since the construction of a concept incorporates our past experience in similar situations and the conceptualization processes recruits associative information to anticipate the likely sensory information to arise, no two instances of a tokened concept will be the same as the current tokened instance will draw upon the previous token instances (Casasanto & Lupyan, 2015). These ad hoc concepts are highly sensitive to current

context and will contain only that information necessary to facilitate our goal-directed cognitive capacities enabling us to navigate our environments and meet the demands of our present tasks (Casasanto & Lupyan, 2015).

An ad hoc conceptual theory stands in stark contrast to traditional psychological theories of concepts. First, by taking each tokened concept to be its own unique construction, the ad hoc conceptual theory rejects the possession of conceptual categories (Casasanto & Lupyan, 2015). There are no conceptual representations stored in long-term memory that may be used to identify stable conceptual categories. Not only are these conceptual categories not shared among individuals, but they are not present across instances of experience. For example, the tokened instance of WRONGNESS in the above moral vignette incorporates information from previous instances of evaluating other states of affairs as ‘wrong’ alongside other information specific to the current instance (Casasanto & Lupyan, 2015). Since this later instance of WRONGNESS will contain information from prior instances, the information being represented by the current ad hoc construction will differ in the evaluative content represented. Ad hoc approaches take conceptual categories to be metacognitive constructions used to group token instances sharing similar properties and enable efficient communication about those instances (Casasanto & Lupyan, 2015). But this similarity should not be mistaken as evidence for conceptual categories defining natural kinds or being necessary posits within our cognitive ontology – any taxonomy emphasizes some features while obscuring others that would undermine their similarity if other measures were prioritized (Casasanto & Lupyan, 2015).

Second, ad hoc concepts reject the distinction between information that constitutes a concept and that information that is associatively activated (Casasanto & Lupyan, 2015, p. 547). In this sense, ad hoc concepts cannot be individuated or distinguished by cores or collections of

information that are statistically, hence incompatible with any notion of conceptual stability or psychological essentialism (Casasanto & Lupyan, 2015). To illustrate this point, consider the concept ROCK that is tokened while searching our visual field in the previous moral vignette. Not all properties of rocks will be represented in this instance, instead, only those properties that are necessary for us to meet our current task-demands. This token instance is significantly different than the token instance of ROCK we construct when sitting in an introductory geology course. In this latter instance, information about the object's composition is important, whereas in the former instance, only the functional properties of being throwable are represented alongside sufficient information to identify the object.

Finally, a fundamental property of ad hoc concepts is their incorporation of information about the expected states of the world, making their constructions predictive rather than passively interpreting sensory information that propagates in a bottom-up manner. This predictive property makes ad hoc concepts “action-oriented” – each conceptual construction aids the agent's navigation through their current internal and external environments in a goal-directed manner. One way to understand how ad hoc concepts are implemented within our neurocognitive architecture is through a generative “forward model” (Barrett & Finlay, 2018, p. 176). These forward models contain multimodal information that serve as predictions of the endogenous and exogenous information encoded by our sensory modalities. These generative forward models enable the action selection, inference formation, and other processes relevant to our task demands in a metabolically and informationally efficient manner through comparing our predictions to the data and resolving errors (Barrett & Finlay, 2018). Errors can be resolved in two different manners: (1) through revising our models to fit the sensory data, or (2) engaging in action to bring our perceived world into congruence with our predictions. The neurocognitive

mechanisms and architecture just discussed is commonly referred to as “Predictive Processing” or “Predictive Coding” (Clark, 2013) and has recently seen substantial interest in the field of cognitive science. But ad hoc concepts do not need to be implemented in this particular framework. Instead, they are compatible with a variety of neurocognitive architectures such as those positing either anatomical or functional modularity, connectionist frameworks, or representationalist versions of dynamical systems.

High-level concepts are comprised of lower-level ad hoc concepts that share a similar construction and conceptualization process (i.e., they can be understood as a generative forward model containing sensory information that issues predictions about the sensory information being encoded within our sensory modalities). This gives rise to a hierarchy of nested concepts that dynamically interact as events unfold, facilitating relevant cognitive processes given our situated context, and issuing behavioral responses to engage within our environments. Since concepts are action-oriented, ad hoc concepts at the highest-levels of the hierarchy will contain within them a variety of possible lower-level action concepts that issue motor commands. Barrett and Finlay (2018) define action concepts as “an integrated summary of multimodal information about motor actions in a particular sensory context” (p. 175). These action concepts are generatively constructed – they pull from our past experience, our successful actions within those experiences, and those actions that are likely to satisfy our task demands given the predicted states of affairs. A high-level ad hoc concept may contain many possible action concepts and the intentional selection of a particular action plan is determined through the dynamic conceptualization process.

Section 4.2: Ad Hoc Moral Concepts

The ad hoc conceptual theory draws heavily upon the embodied cognitive tradition where concepts are instantiated within the modality specific systems themselves (Casasanto & Lupyan, 2015) and the conceptualization process involves reenacting these distributed neurocognitive mechanisms. The kind of information represented during the construction of a concept is multimodal – each conceptual instance contains endogenous and exogenous information that combine to interpret present sensory information, evaluate salient stimuli, and motivate the allocation of attentional resources and processing to select actions or shape our dispositions to select more complex behavioral responses. These properties directly connect ad hoc concepts to their evaluative and motivational features, making this approach a viable candidate to ground a theory of moral concepts. Although a complete characterization of ad hoc moral concepts is beyond the scope of this paper, it will be helpful to briefly consider how this proposal might meet the challenge of motivating situated-action during moral experience outlined in the previous section.

Ad hoc moral concepts are constructed to aid our navigation within social environments and to meet the unique task demands that arise during instances of moral experience. The information represented within a moral concept is multimodal, generative, and predictive – ad hoc moral concepts represent internal and external sensory information by drawing from our past experience, issuing predictions about the likelihood of the incoming sensory information, and priming our dispositions and action plans accordingly. These predictions are not only about the external state of affairs being encoded by sensory modalities, but they are also about our own internal affective states. In this sense, ad hoc moral concepts contain both the descriptive and evaluative components that commonly characterize moral concepts. Ad hoc moral concepts dynamically evolve alongside our environments – Casasanto and Lupyan (2015) refer to this

continuous synchronic integration of contextual information into the conceptualization process as “activation dynamics” (p. 553). The conceptualization process consistently forms new predictions, interprets incoming sensory information, and discrepancies drive the revision of our forward models or the selection of actions.

In the moral vignette introduced in the previous section, the token instance of WRONGNESS results from a moral judgment that negatively evaluates salient stimuli and the predicted outcome of the event which violates our morally relevant expectations. The evaluation is sensitive to the current situation as the content of the concept is comprised primarily by our affective states, the identification of the cat as possessing moral significance, the attribution of moral agency to the children, and more. Our predictions of the likely states of affairs to happen within this experience motivates our particular interpretation and evaluation, which is shaped by our previous interactions in similar situations, acquired moral rules, and moral development. Since ad hoc concepts are action-oriented, moral concepts will also contain within them action plans containing a variety of action concepts that are intentionally selected through a complex process weighting situational constraints against energetic costs (Barrett & Finlay, 2018). On this ad hoc proposal, the incorporation of variable evaluative content and motivational force does not present a challenge to accommodating moral concepts and their roles in motivating situated-action. Instead, the evaluative content is central to the construction of an ad hoc moral concept and the conceptualization process contains error resolution and action selection. In this scenario, WRONGNESS represents the event and our internal affective states, but also motivates the selection of subsequent actions. Given the situation and the dire costs of inaction, the evaluative judgment conferred by tokening WRONGNESS motivates the selection of an action plan to disperse the kids and motor commands to implement our intended behavioral response. Here, ad

hoc moral concepts can accommodate the motivation of situated-action during moral experience without having to appeal to background processes or preselected behavioral responses.

Regardless of the particular Contextualist theory developed, recognizing moral concepts as highly sensitive to the situational context carries consequences for the field of moral cognitive science. The most notable implication concerns the experimental design of moral psychology and its historical reliance on the stimulus-response paradigm. Foundational research in moral cognitive science, such as early work on moral judgment and decision-making, operate under the stimulus-response paradigm – where perception is thought to be carried out in a bottom-up manner and cognitive processing interprets perceived information to produce outcomes. This paradigm influenced the experimental design of this early research and the interpretation of their results. Experimental designs under the stimulus-response paradigm place subjects within highly artificial environments in an attempt to eliminate potentially confounding variables and increase the probability of correctly identifying a valid relationship between a cued stimulus and a subject’s response (Wilford et al., forthcoming). Early moral psychological studies relied upon these experimental designs placing subjects within these artificial environments and either cueing responses to abnormal moral vignettes or crafting specific moral dilemmas called “trolley problems” to assess the formation of moral judgments (Haidt, 2001) and determine the neuroanatomical and functional underpinnings of our moral cognition (Greene et al., 2001).²⁰ However, if the content of a tokened moral concept is sensitive to the agent’s environment, cognitive tasks, and other features of their situational context, then these experimental designs

²⁰ See Gray and Keeney (2015) for a substantive criticism of early moral psychology’s experimental designs.

are inadequate to study moral concepts or other high-level moral phenomena that require the representation of morally relevant information.

One promising alternative to the stimulus-response paradigm, as suggested by Wilford et al. (forthcoming), is to consider perturbation experiments which “aim to identify the precise variable or variables implicated in the ongoing control of a complete activity” (p. 3). On this approach, experimenters would introduce a perturbation condition to manipulate the subject’s behavior either naturally or artificially (Wilford et al., forthcoming). Moral cognitive scientists researching the acquisition and development of moral concepts may find this alternative experimental design to be highly effective when studying how moral concepts shape our dispositions and motivate the selection of situated-action.

Conclusion

Traditional psychological theories of concepts positing invariant representational formats are unable to satisfactorily accommodate moral concepts and their central roles within facilitating our cognitive capacities for moral thought and behavior. Invariantist theories struggle to explain how token instances of moral concepts contain variable evaluative content and motivate the allocation of cognitive resources to produce situationally appropriate and time-sensitive actions. Instances of moral experience containing morally motivated action represent a substantive challenge for these traditional theories and can only be accommodated by positing concepts that possess a flexible representational format. Thus, a satisfactory psychological theory of moral concepts ought to consider Contextualist theories.

References

- Abend, G. (2011). Thick concepts and the moral brain. *European Journal of Sociology*, 52(1), 143–172. <https://doi.org/10.1017/S0003975611000051>
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23. <https://doi.org/10.1093/scan/nsx060>
- Barrett, L. F., & Finlay, B. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in Behavioral Sciences*, 24, 172–179. <https://doi.org/10.1016/j.cobeha.2018.10.001>
- Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364, 1281–1289. <https://doi.org/10.1098/rstb.2008.0319>
- Bloch-Mullins, C. L. (2015). Foundational questions about concepts: Context-sensitivity and embodiment. *Philosophy Compass*, 10(12), 940–952. <https://doi.org/10.1111/phc3.12272>
- Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, 193(5), 1509–1534. <https://doi.org/10.1007/s11229-015-0783-4>
- Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*, 19(4), 371–394. <https://doi.org/10.1177/1088868314566683>
- Casasanto, D., & Lupyan, G. (2015) All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The Conceptual Mind: New directions in the study of concepts* (pp.543–566). MIT Press.
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204.
<https://doi.org/10.1017/S0140525X12000477>
- Crockett, M. J. (2016). How formal models can illuminate mechanisms of moral judgment and decision making. *Current Directions in Psychological Science*, *25*(2), 85–90.
<https://doi.org/10.1177/09637214155624012>
- Cushman, F., Kumar, V., & Railton, P. (2017). Moral learning: Psychological and philosophical perspectives. *Cognition*, *167*, 1–10. <https://doi.org/10.1016/j.cognition.2017.06.008>
- Fingerhut, J., & Prinz, J. J. (2018). Grounding evaluative concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*, 1–7. <https://doi.org/10.1098/rstb.2017.0142>
- Fridland, E. (2017). Skill and motor control: Intelligence all the way down. *Philosophical Studies*, *174*(6), 1539–1560. <https://doi.org/10.1007/s11098-016-0771-7>
- Gantman, A. P., & Van Bavel, J. J. (2015). Moral perception. *Trends in Cognitive Sciences*, *19*(11), 631–633. <https://doi.org/10.1016/j.tics.2015.08.004>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral Foundations theory. In P. Devine & A. Plant (Ed.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Elsevier. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, *6*(8), 859–868.
<https://doi.org/10.1177/1948550615592241>
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. <https://doi.org/10.1080/1047840X.2012.651387>

- Greene, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, 293(5537), 2105–2108. <https://doi.org/10.1126/science.1062872>
- Greene, J. D. (2015). The rise of moral cognition. *Cognition*, 135, 39–42. <https://doi.org/10.1016/j.cognition.2014.11.018>
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834. <https://doi.org/10.1037/0033-295X.108.4.814>
- Harman, G. (1977). *The nature of morality: An introduction to ethics*. Oxford Univ. Press.
- Heyes, C. (2021). Is morality a gadget? Nature, nurture and culture in moral development. *Synthese*, 198(5), 4391–4414. <https://doi.org/10.1007/s11229-019-02348-w>
- Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, 13(1), 1–6. <https://doi.org/10.1016/j.tics.2008.09.006>
- Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science*, 28(3), 280–291. <https://doi.org/10.1177/0963721419831992>
- Kelly, D. (2020). Internalized norms and intrinsic motivations: Are normative motivations psychologically primitive? *Emotion Researcher*, 36–45. <http://emotionresearcher.com/internalized-norms-and-intrinsic-motivations-are-normative-motivations-psychologically-primitive/>
- Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence, (Eds.), *Concepts: Core readings* (pp.3–81). MIT Press.
- Löhr, G. (2017). Abstract concepts, compositionality, and the contextualism-invariantism debate. *Philosophical Psychology*, 30(6), 689–710. <https://doi.org/10.1080/09515089.2017.1296941>
- Machery, E. (2009). *Doing without concepts*. Oxford University Press.

- Margolis, E., & Laurence, S. (2007). The ontology of concepts—Abstract objects or mental representations? *Nous*, *41*(4), 561–593. <https://doi.org/10.1111/j.1468-0068.2007.00663.x>
- May, J. (2019). Précis of *regard for reason in the moral mind*. *Behavioral and Brain Sciences*, *42*, e146. <https://doi.org/10.1017/S0140525X18002108>
- Michel, C. (2020). Concept contextualism through the lens of Predictive Processing. *Philosophical Psychology*, *33*(4), 624–647. <https://doi.org/10.1080/09515089.2020.1742878>
- Nichols, S. (2021). *Rational rules: Towards a theory of moral learning* (First edition). Oxford University Press.
- Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences*, *13*(4), 160–166. <https://doi.org/10.1016/j.tics.2009.01.006>
- Prinz, J. (2006). The emotional basis of moral judgments. *Philosophical Explorations*, *9*(1), 29–43. <https://doi.org/10.1080/13869790500492466>
- Sevinc, G., Gurvit, H., & Spreng, R. N. (2017). Salience network engagement with the detection of morally laden information. *Social Cognitive and Affective Neuroscience*, *12*(7), 1118–1127. <https://doi.org/10.1093/scan/nsx035>
- Sevinc, G., & Spreng, R. N. (2014). Contextual and perceptual brain processes underlying moral cognition: A quantitative meta-analysis of moral reasoning and moral emotions. *PLoS ONE*, *9*(2), e87427. <https://doi.org/10.1371/journal.pone.0087427>
- Sinnott-Armstrong, W., & Wheatley, T. (2014). Are moral judgments unified? *Philosophical Psychology*, *27*(4), 451–474. <https://doi.org/10.1080/09515089.2012.736075>
- Stanford, P. K. (2018). The difference between ice cream and Nazis: Moral externalization and the evolution of human cooperation. *Behavioral and Brain Sciences*, *41*, e95. <https://doi.org/10.1017/S0140525X17001911>

- Tomasello, M. (2016). *A natural history of human morality*. <https://doi.org/10.4159/9780674915855>
- Turiel, E. (2015). Moral development. In R. M. Lerner (Ed.), *Handbook of Child Psychology and Developmental Science* (pp. 1–39). John Wiley & Sons, Inc.
<https://doi.org/10.1002/9781118963418.childpsy113>
- Väyrynen, P. (2011). Thick concepts and variability. *Philosophers' Imprint*, 11(1), 1–17.
- Väyrynen, P. (2021). Thick ethical concepts. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2021). Metaphysics Research Lab, Stanford University.
<https://plato.stanford.edu/archives/spr2021/entries/hick-ethical-concepts/>
- Weiskopf, D. A. (2009). The plurality of concepts. *Synthese*, 169(1), 145–173.
<https://doi.org/10.1007/s11229-008-9340-8>
- Wilford, R., Ardila-Cifuentes, J., Baggs, E., & Anderson, M. L. (2021). *The stimulus-response crisis* [Preprint]. Open Science Framework. <https://doi.org/10.31219/osf.io/rezwc>
- Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, 23(4), 1015–1027. <https://doi.org/10.3758/s13423-015-0948-7>