Kantian Review (2026), **31**, 71–92
doi:10.1017/S1369415426101198

KANTIAN REVIEW
CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Modelling Morality: A Kantian Account of Moral Examples

Saikeerthi Rachavelpula

Philosophy, Columbia University, USA
Email: sr3262@columbia.edu

Abstract

Commentators have recently shown that Kant's view of moral examples is more complex than what appears in the *Groundwork*. Throughout his works, Kant presents explicit reasons for and against the use of moral examples. Reconciling these opposing attitudes has proven difficult. Lacking a satisfactory account of how Kant thinks we should engage with and learn from moral examples, the difficulty has remained. In this paper, I construct and defend an account that resolves the *prima facie* inconsistency by viewing moral examples as analogous to scientific models, thus leading to 'new' and plausible perspectives on recent debates in Kant scholarship.

Keywords: Doctrine of Virtue; moral psychology; models; other minds; chemical analogy

In recent years, commentators have shown that Kant's view of moral examples is far more complex than what appears in the *Groundwork of the Metaphysics of Morals* (*G*).[1] In *G*, Kant is critical with famous claims such as 'one [could not] give worse advice to morality than by wanting to derive it from example' (4: 408). In the *Critique of Practical Reason* (*CPrR*), however, Kant is supportive and advocates for their use throughout moral inquiry. He goes so far as claiming that, if employed properly, examples can make us 'conscious of [our] own freedom' (5: 161). Though such claims are perfectly consistent, they give rise to the following question: *How can Kant's remarks on examples be reconciled into a single account that makes sense of their role in morality?*

In this paper, I construct and defend a novel interpretation of Kant's use of moral examples inspired by a passage in the Doctrine of the Method of *CPrR*. The passage has recently gained commentators' interest concerning Kant's curious claim that morality should adopt a method like those found in science. Taking Kant's advice, I present a view of moral examples as analogous to scientific models. Viewing moral examples as akin to scientific models affords a systematic explanation of their place within Kant's moral philosophy, clarifying both how they might mislead while nevertheless serving as indispensable mediators between the moral law and everyday

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Kantian Review. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

moral life. It additionally offers new and plausible solutions to recent questions of philosophical interest, including interpretative challenges concerning Kant's view of moral justification and issues in contemporary cognitive science concerning theories of 'mind-reading'.

1. Kant on moral examples

Throughout his works, Kant provides several reasons for resisting the use of examples in matters of morality. The first comes from his famous declaration that 'one [could not] give worse advice to morality than by wanting to derive it from example' (*G*: 4: 408). Because moral norms are universal and necessary, they cannot be derived from objects in experience, such as empirical examples. This line of objection is further supported by Kant's point that any example must 'first be appraised in accordance with principles of morality' (4: 408). *A priori* principles of morality are prior to and provide the standard by which we judge empirical examples, so the latter cannot possibly support the former. We can understand this worry as *epistemological* in a formal sense: empirical evidence can never establish a universal and necessary claim because the latter is explanatorily prior to the former.[2] No number of moral examples could serve as evidence for even a single moral principle.

Kant is also sceptical of using examples for *moral* reasons. This can be seen in his remarks on imitation. He writes, 'imitation has no place at all in matters of morality' (*G*: 4: 408). This is because moral acts must be motivated internally by the moral law, not externally by imitation.[3] In imitating another's example, we hinder our own ability to act morally and autonomously. Kant directs this worry at the perfectionist model of the Stoics (Guyer 2016: 268–9). According to Kant, every (normal) human being has the capacity to reason about what to do and to act from her own exercise of reason. Everyone has the capacity for morality, not just the sages. Imitating another denies that capacity.

From these remarks, it might be tempting to conclude that Kant maintains no place for examples in matters of morality.[4] To do so, however, would be a serious mistake (O'Neill 1986; Guyer 1990; Louden 2014). The Doctrine of the Method of *CPrR*, for example, contains extensive remarks on the role of examples in moral education, including their ability to bring the moral law to consciousness. Even in *G*, Kant emphasizes that examples provide '[moral] encouragement' and 'make visible what the practical rule expresses more generally' (4: 409). From such passages, it is clear that Kant thinks examples have some delimited, but nontrivial role in moral inquiry and development. What remains less clear, however, is the mechanism by which this role is fulfilled, given that Kant rules out the obvious evidentiary and imitative functions. Even if we acknowledge *that* examples can make us conscious of our own freedom, we need an explanation of *how* this takes place.

To further complicate matters, any satisfactory account must also explain Kant's qualifications about the kinds of examples that are suitable for moral purposes. It must make sense of Kant's detailed passage warning us to avoid 'noble' (*CPrR*, 5: 157) examples to elicit consciousness of the moral law. He adds that even a suitable example of dutiful action needs to be presented in the right way. It needs to be 'uncontaminated'; otherwise, it too risks leading us astray. Any satisfactory account of

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

moral examples, hence, faces the additional challenge of explaining how and why this function is so easily threatened by 'noble' and contaminated examples.

That Kant returns to the topic of moral examples throughout his writings suggests that they perform some serious and nuanced function in moral inquiry and development. Yet exactly how they accomplish this task is far from obvious. It is this very complexity that makes his treatment of examples challenging to reconstruct and, at the same time, philosophically rich.

2. Kant and models in the philosophy of science

2.1 Kant's Doctrine of the Method

My account relies on an analogy between moral examples and scientific models. This approach is motivated by Kant's remarks in the Doctrine of the Method and recent scholarship in the philosophy of science.

In the Doctrine of the Method, Kant asks how the laws of pure practical reason can be admitted into the human mind and have influence over one's maxims. Kant asks how creatures like us can become aware of the moral law and become accustomed to acting from it. He responds that morality should seek out a method like those found in science. He specifically mentions chemistry (*CPrR*, 5: 163). Though morality is concerned with practical rather than theoretical cognition, Kant thinks it is structurally similar to science. Both are lawful and composed of *a priori* (rational) and *a posteriori* (empirical) components. Both involve the exercise of the faculty of reason. These structural similarities permit a similarity in method.

For Kant, the shared method involves *separation*. In chemical experiments and moral examples alike, experience only provides us with 'impure' phenomena. A phenomenon is impure if its *a priori* and *a posteriori* components are inextricably coupled. In the chemical case, experience only provides composite substances such as 'calcareous earth in hydrochloric acid' as opposed to a pure or isolated 'alkali' (*CPrR*: 5: 92–3). In the moral case, experience only provides an 'empirically conditioned will' in which inclination coexists with the moral law; experience does not provide the isolated moral law. Though separation is not empirically possible, Kant insists one must separate the *a priori* from the *a posteriori* in thought.

Crucially, Kant's method of separation requires us to idealize beyond the objects given in experience. Kant's discussion of regulative ideas in *CPR* expands on the claim. He explains that ideas such as the totality of empirical causes and pure chemical elements ('pure air', 'pure earth', 'pure fire') are 'necessary concepts of reason to which no congruent object can be given in the senses' (A644/B672). We never encounter such a thing as 'pure air' in real life. Nevertheless, 'pure air' *qua regulative idea* plays a necessary role in cognition, which allows us to formulate scientific laws and principles. By idealizing from what is empirically given by positing regulative ideas, one can decouple the *a priori* from the *a posteriori* in thought. Idealization, thus, facilitates separation.

2.2 Models and the philosophy of science

Kant's discussion of regulative ideas has been immensely influential in debates on scientific modelling. Most notably, Hans Vaihinger (1911) argued for a fictionalist

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

view of scientific models based on his interpretation of Kantian regulative ideas. Vaihinger went on to influence the subsequent debate (Bokulich 2009; Fine 1993; Suárez 2009), and there have been recent appeals to Kant's discussion in the literature (Massimi 2017; Spagnesi 2023; Weisberg 2007a).

As these debates suggest, Kant's claims about scientific knowledge are deeply connected to models. In the philosophy of science, modelling is standardly distinguished from other forms of theorizing by two characteristic features: *(1) idealization* and *(2) indirect study*.

Nearly all models exhibit some degree of deliberate idealization. From frictionless planes to point masses, scientists often prefer idealized models to more accurate ones. Their ubiquity and success show that models can give us knowledge about target systems without perfectly representing those systems (Bokulich 2009; Potochnik 2017; Strevens 2004). Still, models need to be representative to some degree; only then can they stand in an analogical relation to their respective target phenomena. The point is that the cognitive value of a model does not depend solely on its power to represent. Accordingly, idealization promotes crucial theoretical virtues such as simplicity, fecundity, and familiarity.

Indirect study further distinguishes modelling. Instead of studying a target system directly via techniques of observation, models are investigated as independent objects of study. Then, conclusions about the model are analogized to generate conclusions about its target system. For instance, instead of studying a hydrogen molecule directly, modellers study the Bohr model. They form conclusions about the electron's orbit in the Bohr model and analogize those conclusions to the experimental behaviour of actual electrons in a given hydrogen molecule.

While idealization figures in many forms of theorizing, what most clearly distinguishes modelling is indirect study (Weisberg 2007b; Hesse 1963). Modelling is neither the application of a pre-given concept to a particular object (e.g., subsuming a particular dog under the universal concept *dog*) nor an exercise in logical deduction (e.g., deducing Kepler's laws from Newton's laws of motion). Rather, modelling is a *reflective* and *exploratory* activity that generates new conceptual connections instead of merely applying existing ones. It enables a reciprocal exchange between abstract structures and empirical phenomena, rendering each more intelligible through the other. In this way, models are said to *mediate* between theory and observation (Morgan and Morrison 1999).

Models are further distinguished by their level of generality. While theory-building seeks universal principles, model-building has a narrower scope, aiming for local rather than global unity. This is evident in nuclear physics, where incompatible models – such as the liquid-drop, shell, cluster, and quark models – each offer coherent yet mutually contradictory accounts. For instance, the liquid-drop model (falsely) treats the nucleus classically and explains fission, while the shell model treats it quantum-mechanically and explains spin. Despite their mutual inconsistency, both are accepted because modelling involves indirect study. We learn from a model only if it is itself a unified and intelligible object of study. Hence, each model must achieve internal coherence even if collective consistency is sacrificed.

The key point is that models differ from other forms of theorizing in their level of generality and in their mediating role. While theories, especially moral ones, have been criticized as 'empty formalisms' (Hegel 1991; O'Neill 1987), models are more

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

localized and enable a bidirectional information flow between abstract principles and particular phenomena.

Although Kant did not know about the contemporary definition or use of models in scientific inquiry, there is a rich history of taking Kant's remarks on separation and regulative ideas as bearing on contemporary debates about scientific models. Moreover, learning from Kantian moral examples involves the two characteristic features of scientific modelling – idealization and indirect study. Suppose we are given in experience an example of dutiful moral action. Kant claims that we learn about our own moral capacity from the example. Already, Kant is calling for indirect study. Namely, we learn about ourselves indirectly by studying another and coming to conclusions about their act. Moreover, because these conclusions concern the moral law, we must idealize beyond the exemplar's empirically conditioned will by positing an idea of the 'pure' moral law. In later sections, I provide a detailed picture of how exactly this process takes place. For now, the key point is that Kant's remarks on moral method show that learning from example involves the two defining features of scientific modelling: idealization and indirect study. This methodological connection grounds an analogy between moral examples and scientific models, dismissing any objections that such a strategy is *ad hoc*.[5]

3. Modelling in science

3.1 Learning from models

This paper's central claim is that Kantian moral examples are analogous to scientific models in an *epistemological* sense. Specifically, it concerns their cognitive role and argues that how we learn from moral examples is analogous to how we learn from scientific models.

Since the publication of Morgan and Morrison's seminal *Models as Mediators* (1999), philosophers of science have made great progress with respect to the latter question. Some even suggest that models give rise to their own style of reasoning dubbed 'model-based reasoning' (Magnani, Nersessian, and Thagard 1999; Nersessian 1999). Contemporary work in the philosophy of science typically begins from a general characterization of models and proceeds to examine how, in virtue of this characterization, models play a distinctive role in scientific reasoning. This paper follows suit. For our purposes, it is not necessary to decide whether models are fictions, set-theoretical structures, or something else entirely. We are interested in *how we learn* from models, where this requires only the assumption that models involve idealization and indirect study. As such, our conclusions are compatible with various metaphysical views of scientific models.

A model's cognitive value arises both from its representational accuracy and from the process of modelling. We draw inferences from a model because it accurately captures the salient features of the target phenomenon, but we also gain insight from a model through the very activity of constructing and manipulating it – by determining what fits together and how different structural assumptions serve scientific aims (Knuuttila and Boon 2011; Cartwright 1983; Woodward 2003).[6]

Consider the emergence of the liquid-drop model of nuclear structure. In their landmark paper, Bohr and Wheeler (1939) modelled it by treating the nucleus as a classical drop of charged fluid. Strictly speaking, this is false: nuclei are quantum

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

systems with discrete energy levels. Yet no quantum model at the time could account for the smooth, collective energy changes observed during fission. Despite its representational inaccuracy, the liquid-drop model was the only one at the time to successfully explain how heavy nuclei split. Aware that this classical assumption was necessary for a coherent model, Bohr and Wheeler deduced that fission occurs at a macroscopic level beyond the quantum scale of individual nucleons. In this way, introducing a representational inaccuracy (a classical nucleus) is exactly what allowed the liquid-drop model to be a unified object of thought that could generate understanding. By constructing and manipulating the liquid-drop model, Bohr and Wheeler engaged in a reflective activity that made them conscious of exactly which assumptions were necessary for an intelligible explanation of nuclear fission.

Modelling is an open-ended, non-algorithmic process in which learning occurs at two stages: construction and manipulation (Morgan 1999). Model construction emphasizes making an abstract structure intelligible and idealizing from the right features of a phenomenon, while model manipulation involves negotiating between that abstract structure and the target phenomenon. Both stages involve a back-and-forth between theory and observation, though each emphasizes one direction more than the other. For our purposes, we consider how learning occurs at both stages.

3.2 A chemical case study

To demonstrate *how* we learn from models via the open-ended activity of modelling, I analyse and illustrate the development of Jean Dumas's famed model of the chemical reaction between alcohol and chlorine. I choose Dumas's case for two reasons: First, Dumas worked shortly after the chemical revolution, within the tradition of chemistry that influenced Kant's remarks on separation; second, Dumas overcame a problem analogous to Kant's formal epistemological worry (Klein 1999; Dumas 1834, 1835, 1840).[7] Namely, he lacked sufficient empirical evidence to establish the causal mechanism driving the reaction. Nevertheless, he overcame this limitation by modelling.

I reconstruct Dumas's modelling as a four-part process adapted from Boumans (1999). This is just one conception of model construction among many. Importantly, different conceptions maintain that these four steps are completed at some point when constructing and manipulating a model.[8] Thus, the key point that we learn from the reflective process of modelling is not unique to this specific account.

3.2.1 *Questioning*. The first step in model construction is formulating a question. This question arises against a background of scientific inquiry and is explanatory in nature. Formulating a question identifies the *explanandum*. This itself is a great achievement, as it involves bringing some empirical phenomenon under concepts that make it cognitively accessible. This original question, though it may be revised, reframes inquiry within a graspable problem-solving context.

In Dumas's case, his question arose from newfound inquiry into the chemical activity of organic substances:

Motivating Question: How does alcohol transform upon the addition of chlorine?

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

At the time, no one could explain what they empirically observed. No one could explain the production of an 'oily' substance which over time transformed into a crystalline substance. In fact, it was impossible to isolate the substance and empirically make out its chemical composition.

3.2.2 Idealizing. Next, one gathers and separates out the *a posteriori* and *a priori* resources to be combined in the model. As mediators, models contain both features. Model-building requires deliberating about which aspects of theory and empirical data to idealize from. The remaining features are then separated with respect to their *a posteriority/a priority*. Given that it is usually the *a priori* that is used to explain the *a posteriori*, separation helps explicitly distinguish the *explanans* from the *explanandum*. This makes clear what is relevant for answering the motivating question and why. In Dumas's case, we have:

A posteriori resources:

(i) Experimental data such as:
   a. The observation that combining a_1 amounts of chlorine and a_2 amounts of alcohol produced a_3 amounts of hydrochloric acid gas and a_4 amounts of a mystery substance.
   b. This mystery substance was initially 'oily' and over d days, it transformed into a crystalline substance.

A priori resources:

(i) *a priori* categories (substance, cause, etc.)
(ii) *a priori* principles of rationality (non-contradiction, *modus ponens,* arithmetic principles, etc.)
(iii) Berzelian formulas (H_2O for the chemical compound of water, etc.)
(iv) Concepts from classical chemistry (chemical compound, chemical reaction, etc.).[9]

3.2.3 Integrating. The third step is to integrate one's resources to create a unified model. Integration is an active process that involves reinterpreting each resource and its relation to the others such that they all cohere and answer to the motivating question. One works back and forth among their various resources until a unified model is achieved; often this involves trial and error. In virtue of this process, the relations between the components of the model become explicit.

For Dumas, this process was far from formulaic (Klein 1999; Dumas 1834, 1835, 1840). After separating his resources, he reinterpreted the transformation as a chemical reaction between two chemical compounds expressed by Berzelian formulas. He set up the following equation:

Chemical reaction (alcohol ($C_8H_{12}O_2$) *and chlorine* (Ch)*)*:
$$C_8H_{12}O_2 + Ch_6 = ChH + x,$$

where x is the mysterious 'oily'-then-crystalline substance.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

Factoring in the observed quantities of each substance, he balanced the equation via basic arithmetic principles. He concluded that the mystery substance was $C_8O_2H_2Ch_6$, meaning that upon the addition of chlorine, each alcohol molecule ($C_8H_{12}O_2$) lost ten atoms of hydrogen (H) and gained six atoms of chlorine (Ch).

Now, Dumas needed to explain why this transformation occurred over several days and produced an intermediary 'oily' substance. After much trial and error, he concluded that it was a *substitution reaction*. This means atoms of hydrogen were withdrawn stepwise and substituted one by one with atoms of chlorine. Because this was his only way to unify all relevant resources and to answer his original motivating question, Dumas knew that he had produced a model that could be analysed as an intelligible object of study.

3.2.4 Inferring. Post-construction, a model can be manipulated to evaluate counterfactuals, make predictions, test theories, design experiments, and so on (Hesse 1966; Morgan 1999). As previously mentioned, constructing models makes explicit the various components of the model and their relations. In virtue of this, models enable inferential reasoning (Suárez 2004, 2009; Hesse 1963).

One way cognitive scientists have understood this is through the notion of *scaffolding* (Larkin and Simon 1987; Sterelny 2004; Zhang 1997). Models can be understood as 'external representation scaffoldings which narrow the space of information search by localizing the relevant features of an object in a perceptually salient and manipulable form' (Knuuttila and Boon 2011: 315). This enables inferential reasoning by making scattered and obscure information systematic and clear. In Dumas's case, he was able to draw powerful inferences, including:

> *Law of Substitution*: Hydrogen atoms (electropositive) can be substituted by chlorine or oxygen atoms (electronegative) in certain organic reactions without any drastic alteration in structure.

Dumas was aware that positing a notion of separation was the only way he could form a unified and coherent model of the chemical reaction. This allowed separation to be an object of thought in its own right. Thus, Dumas could come to conclusions about its scope. In this way, his model became a springboard for further scientific inquiry.

To summarize, Dumas was the first to explain the otherwise puzzling reaction between alcohol and chlorine. By modelling, Dumas overcame a lack of sufficient empirical evidence and developed concepts that proved valuable for later research. All of this was achieved through modelling, which requires determining which resources are relevant, how different components interact, and which assumptions are needed to achieve coherence. Thus, the necessity of certain assumptions becomes explicit, such that the assumptions themselves can be the object of thought, subject to evaluation and refinement.

This extends beyond Dumas's particular case. As seen in Bohr and Wheeler's model of nuclear fission, modelling can illuminate how phenomena at one level – such as quantum processes – affect, or fail to affect, behaviour at another. The same applies to simple 'toy' models like the Bohr model of the atom or the Lotka–Volterra model of predatory-prey relations. In each case, modelling provides a powerful way to organize

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

limited data, uncover the structure of complex phenomena, and recognize the assumptions required for local coherence.

4. Modelling morality

The four-part process of scientific modelling can be fruitfully adapted to moral examples in a way that makes sense of Kant's ambivalence. As in the chemical realm, experience can only provide us with 'impure' phenomena in the moral realm. Experience alone cannot even distinguish a morally motivated act from an immoral one. Kant writes, 'it is absolutely impossible by means of experience to make out with complete certainty a single case in which the maxim of an action otherwise in conformity with duty rested simply on moral grounds and on the representation of one's duty' (*G*, 4: 407). Just as in the chemical case, a model-based approach allows us to overcome this empirical limitation.

I apply the framework of modelling to Kant's example of false testimony in the Doctrine of the Method (*CPrR*, 5: 155–7). He thinks that human beings are born with the moral law within. This means we are born with a predisposition to reason practically according to the moral law, which presents itself in ordinary life when we reflect upon a prospective action. Kant explains that we are conscious of the moral law directly when we formulate a maxim (5: 29–30). He shows his readers that this is the case by presenting them with a striking example. The point of the example is not just to demonstrate a philosophical theory, but to reveal to the readers that they have the moral law within. Providing an example is the first step in a method of moral tutelage that makes readers (or pupils) fully cognizant of the moral law within them and more able to act from it (Guyer 2016: 260).

For Kant, learning from example is not about acquiring new moral principles. It is about bringing *a priori* principles from latency to consciousness. This process reminds us of what we tacitly know from encountering moral issues in life. Though examples might not make us aware of the moral law abstractly as the Formula of Universal Law or the Formula of Humanity, they can prompt us to imagine what would happen if everybody took up a particular action. As Guyer (2016: 264) notes, Kant thinks similarly about learning in mathematics.[10] Though he thinks geometry and arithmetic are (synthetic) *a priori*, he also thinks these subjects need to be taught. Working through the example prompts us to draw on *a priori* principles. As such, examples make *a priori* principles epistemically accessible. Guyer likens this to Plato's *Meno* in which Socrates gets a slave-boy to express geometrical knowledge 'without having been taught but only questioned' (2016: 265).[11]

4.1 Kant's Boleyn man

Kant's example is inspired by the case of Anne Boleyn, who was accused of treason by Henry VIII. In the incident, an unnamed man is pressured to testify falsely against Boleyn. Kant details how the man was bribed with 'large gifts and high rank' before being threatened with defamation, lost friendships, familial disinheritance, and 'unspeakable pain'. Despite this, the man did not testify falsely. He told the truth and 'remain[ed] faithful, without wavering or even doubting' (*CPrR*, 5: 156). Kant claims

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

that even a ten-year-old child would be moved by the example; she would come to admire the man and realize that she too can act morally.

Such claims require Kant's example to do some substantial work, but he leaves it unspecified how the example has such a profound effect on the child. If the example cannot be evidence for moral principles or a call for imitation, how can it cause the child to admire the man and realize that she too can act morally?

Kant, nevertheless, supplies some clues. He remarks that, for the example to work, it must be presented to the child at a stage when she is already able to superficially approve and disapprove of others' actions. At this stage, she can easily judge that the man's act is moral and warrants approval. However, she is unaware of the grounds of her judgement as the moral law within her.[12] Working through the example, Kant claims, builds on the initial superficial judgement and makes her conscious of the moral law. This, in turn, makes her conscious of her own freedom – both her negative freedom not to be determined by inclination and her positive freedom to determine her will autonomously via the moral law. It encourages her by showing her that she too can act morally.

Though Kant supplies these initial clues, it is still a mystery *how* the example makes the child conscious of her own freedom and encourages her. In the following section, I show that the four-part framework of scientific modelling can substantiate Kant's bold claims. The crucial point is that understanding the example requires the child to idealize and engage in indirect study. She cannot directly access the moral law within the man or observe it determining his will. Thus, she must draw upon the moral law within herself and use herself as a model for the man. Moreover, to account for the man's unwaveringness when presented with bribes and threats, she must posit a conception of the 'pure' moral law as the sole source of the man's act.

While it may seem paradoxical that the child uses herself as a model to understand another human being's action, this is not an issue. Kant is clear that the point of the example is to teach the child about herself – namely, that she is free and can act morally. Moreover, this is not the first time Kant calls for using oneself as a model in understanding others. In the Paralogisms (*CPR*, A346/B404-5), Kant says that to represent another mind, we must use ourselves as a model. We cannot have empirical cognition of others' minds, but we are aware of our own rational activity. Aware of our own rational activity, we can project it onto another and understand them as having reasoned like us. Because we cannot have empirical cognition of another's reason, we must use our own. Kant says we are 'entitled' to do this because it is a necessary condition for the possibility of representing another as minded.

4.2 Modelling Kant's Boleyn man

4.2.1 Questioning. Like the scientific case, the moral case begins with a question that arises against a backdrop of inquiry and is explanatory in nature.

Motivating Question: Why did the man refuse to testify falsely?

It arises for the child at a specific stage in her moral development upon confrontation with the example, and it seeks an explanation for some empirical phenomenon – namely, the man's act.[13] Like Dumas, the child identifies an *explanandum* and a

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

question to be answered. This initiates and guides the modelling process within a graspable, problem-solving context.

4.2.2 Idealizing. In the child's case, the relevant *a posteriori* and *a priori* resources come from the empirical details of the example and from the child's own mental capacities.

Child's a posteriori resources:

(i) Details of the example, such as
   a. that the man refused to testify falsely
   b. considerations of pleasure (i.e. various bribes and threats)
   c. her initial superficial judgement of approval of the man's act as moral.

Child's a priori resources:

(i) *a priori* categories (causation, substance, etc.)
(ii) *a priori* principles of rationality (non-contradiction, *modus ponens,* etc.)
(iii) the *a priori* moral law.

For Kant, the moral law is already within us, though we may not be explicitly aware of it. Like the *a priori* category of causation, the moral law is something we bring to experience, which structures our experience. Just as the category of causation automatically shows up for us when confronted with two events that look like necessary succession, the moral law automatically shows up for us when we try to figure out a maxim.[14] In both cases, we are capable of error. Nonetheless, they are necessary conditions for the possibility of theoretical and practical cognition, respectively.[15]

As the child attempts to explain the man's act, she begins to pick out its relevant features. Research shows that children are remarkably skilled at abstracting from the irrelevant features of a problem (Karmiloff-Smith 1996; Starkey and Cooper 1980; Moore et al. 1987). This is because latent principles, such as the principle of non-contradiction, act as attention biases which constrain the hypothesis space entertained for a given problem (Karmiloff-Smith 1996; Thistlethwaite 1951; Hothersall 2004). Though children may not be able to state the principle of non-contradiction explicitly, it implicitly conditions them to avoid contradictory solutions when problem-solving.

This helps us see how the child's resources become involved: some relevant features of the problem, such as the man's refusal to testify falsely, are explicitly picked out; others, such as *modus ponens*, are tacitly evoked for the explanatory task at hand. The latter class of resources is not explicit, but operates as a set of constraints on the space of possible solutions. As the child integrates her resources, they can be brought to consciousness.

4.2.3 Integrating. To see how the child integrates her resources, we need to understand how the moral law shows up for her when modelling. As she tries to figure out a possible maxim for the imagined case, she sees that considerations of pleasure do not explain the man's act. Kant's elaborate detailing of the bribes and threats faced by him helps her realize that contingent details about the man and his specific situation cannot be used to explain his act. Hence, another strategy is required. In

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

searching for another explanation, she considers what she would do in the man's situation.

This allows the child to deliberate about the man's case from a first-person standpoint, where, as Guyer (2016: 260) explains, the moral law 'reveal[s] itself in reflection upon prospective action'. To figure out how she would act, her latent possession of the moral law leads her to imagine what would happen if everyone testified falsely. Recognizing that this would yield a contradiction, she sees that such a maxim cannot be coherently universalized or willed as a law for herself. From this, she forms the practical judgement, 'I cannot testify falsely'. In reflecting on herself in this way, she implicitly appeals to the moral law. Her conclusion is not merely theoretical but practical – it expresses an intention grounded in the moral law.

According to Kant, this entire line of reasoning is *self-conscious*. We are automatically aware of our own rational activity. Thus, the child's entire deliberation about what she would do and whether her action could be willed universally is brought to consciousness. She is aware that she cannot testify falsely because such a maxim cannot be willed as a universal law. In this sense, she becomes aware of the moral law both as the principle that determines her conclusion and as the ground of her intention. Even without explicitly articulating the moral law, her practical reflection brings it to consciousness as the guiding principle of her will.

The crucial points are: (1) in reasoning through the example, the moral law is tacitly evoked in her deliberation, and (2) because this reasoning is self-conscious, it renders her aware of the moral law itself. Notice, these claims do not depend on any particular interpretation of Kantian practical reason; they follow from Kant's own claims.

During this process, the child acknowledges she is using herself as a model. That is why it is formulated in the first-person 'I'. At this point, she does not yet possess a model of the man's act. She must *project* or generalize her line of reasoning onto the man. She must further reason that, given her resources, the only way to explain his act is as having reasoned like her. She combines this with her other resources to finally construct a model of the man's act with the moral law as its determining ground, despite pressures of pleasure to do otherwise.

Still, one might object that the child may not seamlessly generalize her own thinking to the man. After all, the two are very different. Why must she reason that the man acted according to the moral law? Why can she not reason that he acted from some other motive, such as his desire to impress others? This objection is addressed by Kant's clue that the child needs to have superficially approved of the man's act from the outset. To explain her initial approval, she *must* reason that the man acted morally. Otherwise, he would not have acted in a way that warrants her initial approval. To put it simply, given her resources, she has no other option but to use the moral law if she is to understand the man's act. In this way, her initial approval is necessary for her reasoning that the man acted from the moral law.

Thus, the child has successfully constructed a model of the man's act. She has integrated her resources such that they all cohere and some provide explanatory grounds for others. Namely, the moral law provides explanatory grounds for the man's particular act.[16] Just as Dumas explains the chemical reaction between alcohol and chlorine as a substitution reaction, the child explains the man's refusal to testify falsely as an act motivated by the moral law. Integration makes clear the rational

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

dependence between the components in a model. In Dumas's case, he saw how his *a priori* and *a posteriori* resources work together to drive a chemical reaction. In the child's case, she saw how the moral law is separate from and authoritative over the sum-total of our sensible inclinations. In both instances, the modellers uncover the rational relations between their resources, whatever the details of that relation may be.

4.2.4 Inferring. From here, the model becomes grounds for inference. Constructing a model makes explicit its constituent components and their relations. This, in turn, enables inferential reasoning. This feature of modelling, I will now argue, can support Kant's substantial claims that the example makes the child conscious of her own freedom and that she too can act morally.

From constructing the model, the child is aware of the moral law as the determining grounds of the man's act – an act she judges to be moral. From this position, she can form inferences. For instance, she can infer that the man would have been able to act morally in all his actions, not just the one described. Crucially, she can also make inferences about the moral law within herself. Since the child understood the man as having reasoned like her, she became aware that the very ingredient that allowed the man to act morally – the moral law – is already present in her as the standard she used to approve of his act in the first place. She realizes that it was not a contingent quirk about the man that gave rise to his act. Rather, it was the moral law within – something she is now conscious of possessing herself. As such, she infers that she too can act morally.

Kant admits that his method is not foolproof (*CPrR*, 5: 153). Though examples enable the child to make the inference and they may be our best strategy, they cannot guarantee it. There is always the possibility of failure.[17] Nonetheless, there is also the possibility of success. And in cases of success, examples make the child conscious of her own freedom. In becoming conscious of the moral law within herself, she realizes that she too is not determined by inclination and can determine her will autonomously via the moral law. In other words, she realizes she is free. She realizes that she is not just a spectator of morality, one who judges from a distance. She is also a participant, one who is capable of moral action and who is to be judged as well.

We can now present a comprehensive explanation of how the example is meant to encourage the child: to understand it, she constructs a model of it based on herself. This triggers the rational activity of deliberation, whereby she becomes conscious of the moral law. By projecting this line of reasoning onto the man, she constructs a unified model that explains the exemplar's behaviour. Conscious of the model's constitutive elements and their relations, she draws inferences from it. Crucially, she realizes that what allowed the exemplar to act morally (the moral law) is already within herself. Thus, she infers that she too can act morally.

In this way, we can learn from examples via the activity of modelling.[18] The example is not evidence for the moral law nor is it something to be imitated. The example presents a graspable problem-solving context in which the moral law must be evoked to generate a solution. To extend the analogy of a chemical reaction, the example is merely a *catalyst*. What we really learn from is the modelling process it initiates. We learn from the self-conscious activity of reasoning with the

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

moral law when modelling. Ultimately, it is reason itself that delivers consciousness of the moral law.

Kant's insight into the self-consciousness of reason goes beyond the cognitive-scientific notion of scaffolding. While scaffolding explains that models highlight salient features of a target phenomenon, it does not explain how. Our attention can be directed in a myriad of ways. Why is modelling special? Kant's point that reasoning is self-conscious answers this question. Modelling makes explicit a model's constitutive elements and the relations between them because it explicitly involves reasoning about which components to idealize from and how these components can be made to cohere with one another. Because this rational activity is self-conscious, we become aware of the rational grounds of the conclusions we draw. Self-consciousness reveals why certain resources are relevant and why specific relations between them hold. By modelling, we are *automatically* aware of these considerations. Thus, the latent principles and assumptions necessary to achieve such conclusions are brought to consciousness.

4.3 Noble and uncontaminated examples

Any account of Kantian moral examples needs to explain why only certain examples can make us conscious of the moral law. My account has the resources to understand the rationale behind his claim and Kant's own use of examples throughout his texts. In the Doctrine of the Method, Kant warns against the use of 'noble' (*CPrR,* 5: 157) or supererogatory examples. The key to making sense of this claim is recalling that we learn from models in virtue of their evocation of the moral law. In cases of supererogatory actions, the determining ground is not the moral law, but some other factor such as 'feeling' (ibid.). This prevents the modelling process from evoking the moral law, which prevents our ability to learn from the example. For this reason, it is important that our examples avoid depicting overly sentimental, supererogatory actions.[19]

Additionally, my account implies that we should present models that possess theoretical virtues such as familiarity, fecundity, and simplicity.[20] This allows us to understand Kant's claim that moral examples must be 'pure' and 'uncontaminated'. An example is *pure* if it is not contaminated by experience. It must present the moral law as the sole determining ground of action. Purity makes clear what the moral law can accomplish on its own, and this advances the systematicity of cognition just like the theoretical virtue of simplicity.[21] Understanding purity as a theoretical virtue, we can see why Kant presents cases in which duty and inclination so strongly diverge. Providing an *explanans* for a *pure explanandum* necessitates the evocation of the moral law. No other explanation suffices. While some have mistakenly seen this as Kant's moralistic tendency to make morality seem as difficult as possible, we can understand it as Kant's way of avoiding contaminated examples where inclination and duty converge. Kant avoids contamination to further the systematicity of cognition. In demonstrating what the moral law can accomplish on its own, he also encourages pupils to believe in their own moral strength.

Finally, my account makes sense of Kant's own usage of examples throughout his writings. Despite his critical remarks on examples, Kant utilizes many examples throughout his works. Perhaps the most famous is the shopkeeper example from the

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

G, which has been criticized on the grounds that it is under-described (Jensen 1989). In my view, Kant's under-description is an advantage rather than a shortcoming. The incompleteness and underdetermination of empirical data necessitate the evocation of the moral law for the explanatory task at hand. It also gives us a head start in abstracting from irrelevant particulars.[22] Because the way we learn from examples comes from our consciousness of the moral law, it makes sense for Kant to avoid clouding the picture with superfluous empirical details.

4.4 Disanalogies

While the account I provide emphasizes the similarities between a scientist like Dumas and a moral agent like the child, it is important to keep in mind their disanalogies. Two notable disanalogies need to be addressed concerning the sophistication of scientists and the role of emotion. In each case, I argue that the disanalogy does not undermine the account.

The first disanalogy is a familiar one in the philosophy of science. Dumas, a trained chemist, engages in modelling with a degree of sophistication that far exceeds that of the child. It may therefore seem that a model-based account risks over-intellectualizing the child's activity. This concern, however, rests on a misunderstanding of the level of generality at which the analogy is drawn. The point is not that the child engages in the same technical practices as the scientist, but that her reasoning exhibits the same *form* of reflective and exploratory engagement characteristic of modelling, as distinct from mere concept application, deduction, or imitation. The four-part structure of modelling invoked in Dumas's case could equally well be illustrated by toy scientific models, such as the Lotka–Volterra model of predator-prey dynamics taught to school children. In all cases, the key point is that learning occurs in the same form via the construction and manipulation of a model. Differences in instrumental or mathematical sophistication are too fine-grained to undermine the analogy at this level of generality.

A similar point can be used to address the role of emotion. At first, it may seem that emotion plays a much larger role in morality than science. However, as important work in the feminist philosophy of science has shown, affective dispositions affect judgements of salience in scientific inquiry. They affect what *a priori* and *a posteriori* come to mind when modelling. Similar claims apply in the moral case. In both cases, modelling allows for emotion to play a nontrivial, supportive role. Of course, it may play other roles. The point is that the role of emotion, however large or small, is compatible with modelling in either case. Given the contentious nature of the role of emotion in Kantian ethics (Grenberg 2013; Ware 2015), it is a virtue of the modelling account that it does not come with a particular view of emotions baked in.

5. A springboard for further inquiry

A model-based account shows how Kant can substantiate his bold claims that moral examples have an essential place in morality. This account drew on complex debates in the philosophy of science and Kantian practical reason. What is the point of going to such trouble? In this section, I show that there is a payoff. I show that it yields 'new' and plausible perspectives on two debates of contemporary interest in Kant

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

scholarship and cognitive psychology: Kant's views on moral justification and the human ability to know other minds.

5.1 Experiments and moral justification

Recently, there has been some interest in how the sciences influence Kant's view of morality (Ameriks 2000; Rosenkoetter 2011; Sticker 2017). This is due to Kant's remarks on chemical experiments as a paradigm for practical philosophy. For historical context, Kant witnessed the Chemical Revolution spearheaded by Lavoisier. Fully aware of the revolution's advances, he saw empirical experiments as playing a unique role in chemistry that morality should emulate.

As mentioned, both morality and science can benefit from a method of separation. In the Doctrine of the Fact of Reason (*CPrR*, 5: 31–2), Kant furthers the similarity. He states that a moral philosopher can establish another's consciousness of the moral law by setting up an experiment akin to a scientist setting up a chemical experiment. Willascheck (1992) was the first to read this as Kant's call for moral philosophers to employ thought experiments, interpreting the Fact of Reason as a kind of 'act' or 'deed'. Since Willascheck's influential reading, philosophers have noted that calling for thought experiments has interesting implications for Kant's thinking on moral justification.

Ware (2014, 2021), for instance, utilizes Willascheck's reading to deduce that moral justification takes place from the first-personal, common standpoint of ordinary life as opposed to the third-personal standpoint of philosophical speculation. Following Willascheck, Ware takes Kant to present a thought experiment in the Fact of Reason when presenting the example of the Boleyn man. Imagining the hypothetical scenario makes the reader conscious of the moral law and of her own freedom. As Ware explains, the thought experiment demonstrates 'the way we separate duty from happiness and acknowledge the authority of the former'.[23] After arguing for his reading, Ware notes there remains an unresolved inconsistency: Why does Kant treat the reader as a passive spectator of the Fact of Reason while also seeming to call for the moral law from a practical, first-person perspective?[24]

One contribution of my account is that it resolves this inconsistency. We can understand the thought experiment as presenting an example that initiates the self-conscious activity of modelling. Crucially, this involves using ourselves as a model. Exactly because we cannot directly observe another's rational activity, it involves reasoning from our own point of view and projecting that line of reasoning onto another. This explains how we can make sense of another's action while reasoning from the first-person perspective of an active participant. Modelling involves recognizing that we are using ourselves as a model. That is why we project our line of reasoning onto another and combine it with our other resources to construct a coherent model. As such, learning from examples involves a participatory dimension despite not bringing us into the narrative scene. The universality of moral principles enables us to learn about another's motive and about our own moral capacity.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

5.2 Other minds and 'mind-reading'

A model-based account shows how we can learn from moral examples by using ourselves as a model. This is, of course, not the first time Kant suggests using ourselves as a model for another. In the Paralogisms, Kant claims we must model to understand another as minded. He justifies this move in his statement that 'we must necessarily ascribe to things *a priori* all of the properties that make up the condition under which alone we think them' (*CPR*, A346/B404-5). To understand another human being as minded, we must ascribe to them *a priori* the properties that constitute being minded. However, the only way to uncover these properties is from within. We observe in ourselves the qualities of being minded (e.g., that a mind understands its states as connected), and we project these qualities onto the other. We are entitled to do this because it is the only way we can represent them as minded.

Kitcher (2011: ch. 15 ['Kant our Contemporary']) points out that Kant's view has the potential to fill a gap in current theories of how humans can understand other minds. In the last few decades, these debates have centred around questions about 'mind-reading': how one knows the content of another mind's thought.[25] In contrast, Kant's discussion in the Paralogisms is about how we can represent another mind *as such*. My model-based view extends Kant's discussion to contemporary cases of mind-reading in which we attempt to know the content of another's mind. The subtle distinction allows it to overcome limitations suffered by the two leading theories of mind-reading.

The two leading accounts of children's mind-reading – the theory-theory (Gopnik, Meltzoff, and Kuhl 1999) and the simulation theory (Gordon 1986; Goldman 2006) – are distinguished by how they explain the Maxi Experiment. In the classic experiment, a child sees a puppet, Maxi, place a chocolate in one location. Then, the child observes the chocolate being moved without Maxi's knowledge; the child is then asked where Maxi will look for it. According to the theory-theory, children possess primitive psychological theories that guide their inferences about beliefs and desires; this enables them to predict that Maxi will look where she last left the chocolate rather than where it is actually located. Simulation theorists, by contrast, argue that children do not require such theoretical knowledge; they simply imagine themselves in Maxi's situation and reason about what they would do.

While both accounts capture aspects of children's reasoning, they face limitations. The theory-theory struggles to explain the role of experience and education if children already come equipped with robust theories, whereas the simulation theory cannot easily handle cases in which the other's situation is dissimilar from our own, such as imagining the pains of labour or perspectives very unlike our own.

A Kantian model-based account resolves these limitations by combining insights from both approaches. Rather than positing innate theories, it holds that humans are born with minimal, latent categories and principles of rationality that allow them to model others as minded and to generate hypotheses about their reasoning. In familiar cases, the child may project or apply their first-personal model. In dissimilar or complex situations, however, alternative hypotheses can be generated and evaluated against one another, much like how scientists test rival models. This approach explains how children actively reason about others without requiring them to simulate every possible perspective or rely on pre-formed theoretical knowledge.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

Learning, then, is not passive manifestation or simulation but active engagement with examples that elicit reasoning according to latent principles.

The Paralogisms demonstrates how we can know another mind as such. My model-based theory demonstrates how we can know the particular content of another mind. By using ourselves as a model to understand another, we recognize that others enjoy the property of mindedness and are capable of thinking just like us. We also recognize that in any particular instance, they do not have to think like us. The innate scaffolding of the mind allows us to develop alternative hypotheses. In particular cases, we evaluate rival hypotheses with respect to their ability to coherently account for all relevant details.

My model-based view demonstrates in detail how we model the particular content of another by combining empirical details with *a priori* categories and principles. Uncovering this mechanism reveals how abstract principles make contact with particular situations, how latent principles can become known, and why modelling other minds is beneficial for tasks like moral education. As such, my model-based theory has the resources to extend Kant's theory of practical reason to a theory of moral development. For Kant, children are neither scientists nor simulation machines. They are modellers.

6. Conclusion

A modelling account understands the epistemic value of moral examples as stemming from the self-conscious rational activity of modelling. This affords a systematic explanation of their place within Kant's moral philosophy, clarifying both how they might mislead while nevertheless serving as indispensable mediators between the moral law and everyday moral life. Moreover, it has significant payoff for other debates concerning Kant's discussion of moral justification and even cognitive psychological debates about 'mind-reading'. A modelling view is promising for other contemporary debates as well, such as those concerning the role of thought experiments in metaphysics and the role of emotions in morality. Ultimately, when it comes to learning from example, there is a spirit of 'learning by doing' reminiscent of Kant's imploration to think for oneself in his 1784 essay 'What is Enlightenment?'.

Acknowledgements. I am very grateful to Patricia Kitcher for her valuable feedback and support. Additional thanks to Aaron Friedman, Joe Hamilton, Justin Clarke-Doane, Katja Vogt, Martin Sticker, SWIP-NYC, and two anonymous referees.

Notes

1 Citations appear with abbreviation, followed by volume/page number from the Akademie Ausgabe, *Kants gesammelte Schriften*, edited by Königlich Preussische Akademie der Wissenschaften, and successors (29 vols.) (Berlin: de Gruyter, and predecessors, 1900-). The only exceptions to this rule are in quotations from the *Critique of Pure Reason*, where I follow the standard practice of referring to the first edition of 1781 (A) and the second edition of 1787 (B). Translations are from Kant (1996, 2007, 2010, 2012, 2019). Abbreviations: *G, Groundwork of the Metaphysics of Morals*; *CPR, Critique of Pure Reason*; *CPrR, Critique of Practical Reason; R, Religion Within the Boundaries of Mere Reason.*

2 Compare this to an Aristotelian picture, where the *phronimos* sets the standard for moral virtue. On this view, one can build a comprehensive moral theory by investigating the *phronimos* as an exemplar.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

3 In accepting this worry, my account does not include Kant's discussion of *Beispiel*. Even in the case of *Beispiel*, however, Kant warns against the extended use of examples for imitative purposes. See Kant's Lectures on Pedagogy (9: 461–6; tr. R. B. Louden in Kant 2007) and Guyer (2016).
4 Peterson et al. (2010: 141) write: 'Kant's major works on ethics do not place even minimal consideration on moral exemplarity, nor do they give any significant reference to moral exemplars in the course of the arguments put forth'. Similarly, Nussbaum (1986: 4–7) argues that the 'enormous influence of Kantian ethics on our intellectual culture' has prevented us from understanding ancient ethics and particularly the exemplar-centred notion of practical deliberation given by Aristotle.
5 Spagnesi (2023) further shows how this underexplored, yet fruitful line of inquiry can be utilized to solve puzzles about idealization in the philosophy of science.
6 Recall Cartwright's (1983) famous statement that 'the truth doesn't explain much'.
7 Not to mention, Dumas's activity of modelling is well-documented thanks to his rigorous journaling, and his case is easily understood even for those lacking specialized knowledge in chemistry.
8 For instance, Knuuttila and Boon (2011) and Morgan and Morrison (1999) agree that constructing models involves the general steps of questioning, idealizing, integrating, and inferring.
9 Though *a priori* principles of rationality and *a priori* categories may not be consciously gathered like Berzelian formulas, I include them because they are utilized during model construction and can be made explicit from modelling.
10 Guyer (2016: 264) writes: 'Kant hardly assumes that because geometry and arithmetic, which describe the structure of space and time, are *a priori*, they do not need to be taught. They do need to be taught, and are taught by the use of examples – the construction of figures, the addition of sums – that illustrate their general principles. The same is true in the case of the general principle and concepts of morality, even though they are a priori'.
11 Of course, Kant does not think 'finding knowledge within oneself' is a matter of recollection as Socrates does. However, he agrees that we can 'find knowledge within ourselves' by bringing what is *a priori* to consciousness via examples. Guyer (2016) argues convincingly for this claim.
12 Kant thinks children can easily and reliably identify an act as moral or immoral with very minimal training. This training involves a game where children are scored based on their ability to successfully classify various examples as moral/immoral acts. See *CPrR*, 5: 153–6.
13 Developmental psychologist Karmiloff-Smith (1992) observes that children at a certain age are no longer satisfied with successfully judging an action to be right or wrong. They want to know *why*. Such questions naturally arise at a specific age in their development.
14 Kant thinks we become immediately conscious of the moral law 'as soon as we draw up maxims of the will for ourselves' (*CPrR*, 5: 29).
15 In *Religion within the Boundaries of Mere Reason* (6: 26–8), Kant labels this our inextricable predisposition to humanity – our natural receptivity to the moral law and capacity to be moral beings.
16 This implies that only those who already possess the moral law within can learn from examples. If one does not already possess it, then it does not arise in the activity of modelling, and one cannot become conscious of it. This coheres with Kant's assumption that all (normal) human beings possess the moral law within and that if they do not, there is nothing you can do for them. See *Religion*, 6: 26, note a).
17 Note that the possibility of failure is present in other forms of learning, so this is not a unique problem for the moral case. Moreover, it need not be a bad thing. It is a consequence of the fact that successfully learning from an example is an autonomous achievement that cannot be spoon-fed.
18 See Kitcher (2017), Willaschek (1992), who generalize the account to Kant's Fact of Reason.
19 Guyer (2016) claims that Kant would necessarily reject examples from fiction. If we understand Kant's avoiding supererogatory examples because they do not necessitate the invocation of the moral law, then Guyer is clearly mistaken. Indeed, an example from 'real life' could be fictional in having been motivated by a covert impulse of self-love. However, we can still learn from fictional examples so long as they are plausible and not supererogatory. A plausible, dutiful example from fiction would be much preferred to a supererogatory action from 'real life'. This coheres with the familiar point that models in science can be useful despite being clearly false.
20 Kant's own examples tend to be simple and familiar. Moreover, he refers to the Boleyn Man many times throughout the *CPrR*, a testament to its fecundity.
21 See Guyer (1990) for an excellent discussion of the role of purity in cognition.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

22 Empirical research (Bryant 1974) shows that children exhibit more success in problem-solving when irrelevant features have already been abstracted from.
23 It should not surprise us that this is the same example Kant prompts us to present to children in the Doctrine of the Method. In both cases, the aim is the same: to make the reader and/or child conscious of the moral law.
24 Ware (2021: 60) discusses how Kant prompts us to engage in a thought experiment in the *CPR* by having us imagine removing from a body 'everything that is empirical in it – the color, the hardness or softness, the weight, even the impenetrability'.
25 See Astington, Harris, and Olson (1988) for an excellent discussion of children's theories of the mind.

References

Ameriks, K. (2000), *Kant's Theory of Mind: An Analysis of the Paralogisms of Pure Reason*. Oxford: Oxford University Press. Online edn, Oxford Academic, 1 Nov. 2003. https://doi.org/10.1093/0198238975.001.0001. Accessed 12 Febuary 2026.
Astington, J.W., Harris, P.L. and Olson, D.R. (eds.) (1988) *Developing Theories of Mind*. Cambridge: Cambridge University Press.
Bohr, N.H. and Wheeler, J.A. (1939) 'The mechanism of nuclear fission'. *Physical Review*, 56, 426–50.
Bokulich, A. (2009) 'Explanatory fictions'. In M. Suárez (ed.), *Fictions in Science: Philosophical Essays on Modeling and Idealization* (New York and London: Routledge), 91–109.
Boumans, M. (1999) 'Built-in justification'. In M.S. Morgan and M. Morrison (eds.), *Models as Mediators. Perspectives on Natural and Social Science* (Cambridge: Cambridge University Press), 66–96.
Bryant, P. (1974) *Perception and Understanding in Young Children: An Experimental Approach* (1st ed.). Routledge. https://doi.org/10.4324/9781315534251.
Cartwright, N. (1983) *How the Laws of Physics Lie*. Oxford: Oxford University Press.
Dumas, J. (1834) 'Recherches de chimie organique'. *Annals de Chimie et de Physique*, 56, 112–50
—— (1835) *Traité de Chimie Appliquée aux Arts*. 8 vols. (1828-46), vol. 5. Paris: Béchet Jeune.
—— (1840) 'Mémoire sur la loi des substitutions et la théorie des types'. *Comptes Rendus*, 10, 149–78.
Fine, A. (1993) 'Fictionalism'. *Midwest Studies in Philosophy*, 18, 1–18.
Goldman, A.I. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
Gopnik, A., Meltzoff, A.N. and Kuhl, P.K. (1999) *The Scientist in the Crib: Minds, Brains, and How Children Learn*. New York: William Morrow & Co.
Gordon, R.M. (1986) 'Folk psychology as simulation'. *Mind and Language*, 1(2), 158–71.
Grenberg, J. (2013) *Kant's Defense of Common Moral Experience: A Phenomenological Account*. New York: Cambridge University Press.
Guyer, P. (1990) 'Reason and reflective judgement: Kant on the significance of systematicity'. *Noûs*, 24(1), 17–43.
—— (2016) 'Examples of moral possibility'. In P. Guyer, *The Virtues of Freedom: Selected Essays on Kant* (Oxford: Oxford University Press), 260–72.
Hegel, G.W.F. (1991) *Hegel: Elements of the Philosophy of Right*. Trans. H.B. Nisbet, A. W. Wood (ed.), Cambridge: Cambridge University Press.
Hesse, M. B. (1963) *Models and Analogies in Science*. Notre Dame, IN: University of Notre Dame Press.
Hothersall D. (2004) *History of Psychology*. 4th ed. Boston: McGraw-Hill.
Jensen, H. (1989) 'Kant and moral integrity'. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 57(2), 193–205.
Kant, I. (1996) *Critique of Pure Reason*. With Introduction by P. Kitcher, trans. W. S. Pluhar. Indianapolis, IN: Hackett Publishing Co.
—— (2007) *Anthropology, History, and Education*. Ed. G. Zöller and R.B Louden. Cambridge: Cambridge University Press.
—— (2010) *Critique of Practical Reason*. With Introduction by S. Engstrom, trans. W. S. Pluhar. Indianapolis, IN: Hackett Publishing Co.
—— (2012) *Groundwork of the Metaphysics of Morals*. With Introduction by C. Korsgaard, trans. and ed. M. Gregor and J. Timmermann. Cambridge: Cambridge University Press.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

—— (2019) *Religion within the Boundaries of Mere Reason and Other Writings*. With Introduction by R.M. Adams, trans. and ed. A. W. Wood and G. Di Giovanni. Cambridge: Cambridge University Press.
Karmiloff-Smith, A. (1996) *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Boston, MA: MIT Press.
Kitcher, P. (2011) *Kant's Thinker*. Oxford: Oxford University Press.
—— (2017) 'A Kantian argument for the formula of humanity'. *Kant-Studien*, 108(2), 218–46.
Klein, U. (1999) 'Techniques of modelling and paper-tools in classical chemistry'. In M.S. Morgan and M. Morrison (eds.), *Models as Mediators. Perspectives on Natural and Social Science* (Cambridge: Cambridge University Press), 146–7.
Knuuttila, T. and Boon, M. (2011) 'How do models give us knowledge? The case of Carnot's ideal heart engine'. *European Journal for Philosophy of Science*, 1(3), 309–34.
Larkin, J.H. and Simon, H.A. (1987) 'Why a diagram is (sometimes) worth 10,000 words'. *Cognitive Science*, 11, 65–99.
Louden, R. (2014) 'Making the law visible: the role of examples in Kant's ethics'. In R. Louden, *Kant's Human Being: Essays on His Theory of Human Nature* (Oxford: Oxford University Press), 91–104.
Magnani, L., Nersessian, N.J., and Thagard, P. (eds.) (1999) *Model-based Reasoning in Scientific Discovery*. Dordrecht: Kluwer Academic Publishers.
Massimi, M. (2017) 'What is this thing called "scientific knowledge"? Kant on imaginary standpoints and the regulative role of reason'. *Kant Yearbook*, 9, 63–84.
Moore, D., Benenson, J., Reznick, S., Peterson, M., and Kagan, J. (1987) 'Effect of auditory numerical information on infants' looking behavior: contradictory evidence'. *Developmental Psychology*, 23, 665–70.
Morgan, M. (1999) 'Learning from models'. In M.S. Morgan and M. Morrison (eds.), *Models as Mediators. Perspectives on Natural and Social Science* (Cambridge: Cambridge University Press), 347–88.
Morrison, M. and Morgan, M.S. (1999) 'Models as mediating instruments'. In M.S. Morgan and M. Morrison (eds.), *Models as Mediators. Perspectives on Natural and Social Science* (Cambridge: Cambridge University Press), 10–37.
Nersessian, N.J. (1999) 'Model-based reasoning in conceptual change'. In L. Magnani, N.J. Nersessian, and P. Thagard (eds.), *Model-based Reasoning in Scientific Discovery* (Dordrcht: Kluwer Academic Publishers), 5–22.
Nussbaum, M. (1986) *The Fragility of Goodness. Luck and Ethics in Greek Tragedy and Philosophy*. Cambridge: Cambridge University Press.
O'Neill, O. (1986) 'The power of example'. *Philosophy*, 61(235), 5–29.
—— (1987) 'Abstraction, idealization and ideology in ethics'. *Royal Institute of Philosophy Lecture Series*, 22, 55–69.
Peterson, G.R., Spezio, M., Van Slyke, J.A., Reimer, K., and Brown, W. (2010) 'The rationality of ultimate concern: moral exemplars, theological ethics, and the science of moral cognition'. *Theology and Science*, 8(2), 139–61.
Potochnik, A. (2017) *Idealization and the Aims of Science*. Chicago, IL: University of Chicago Press.
Rosenkoetter, T. (2011) 'Kant on construction, apriority, and the moral relevance of universalization'. *British Journal for the History of Philosophy*, 19(6), 1143–74.
Spagnesi, L. (2023) 'Regulative idealization: a Kantian approach to idealized models'. *Studies in History and Philosophy of Science*, 99(C), 1–9.
Starkey P. and Cooper R.G. (1980) 'Perception of numbers by human infants'. *American Association for the Advancement of Science*, 210(4473), 1033–5.
Sterelny, K. (2004) 'Externalism, epistemic artefacts and the extended mind'. In R. Schantz (ed.), *The Externalist Challenge. New Studies on Cognition and Intentionality* (Berlin and New York: de Gruyter), 239–54.
Sticker, M. (2017) 'Experiments in ethics? Kant on chemistry, means of education and methods of practical philosophy'. *Idealistic Studies*, 46(1), 41–63.
Strevens, M. (2004) 'The causal and unification approaches to explanation unified—causally'. *Noûs*, 38(1), 154–76.
Suárez, M. (2004) 'An inferential conception of scientific representation'. *Philosophy of Science*, 71(5), 767–79.
—— (ed.) (2009) *Fictions in Science: Philosophical Essays on Modeling and Idealization*. New York: Routledge.

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press

Thistlethwaite D. (1951) 'A critical review of latent learning and related experiments'. *American Psychological Association Psychological Bulletin*, 48(2), 97–129.
Vaihinger, H. (1911) *Die Philosophie des Als Ob*. Berlin: Verlag von Reuther & Reichard.
Ware, O. (2014) 'Rethinking Kant's fact of reason'. *Philosophers' Imprint*, 14(32).
—— (2015) 'Accessing the moral law through feeling'. *Kantian Review*, 20(2), 301–11.
—— (2021) *Kant's Justification of Ethics*. Oxford: Oxford University Press.
Weisberg, M. (2007a) 'Three kinds of idealization'. *The Journal of Philosophy*, 104(12), 639–59.
—— (2007b) 'Who is a modeler?' *British Journal for the Philosophy of Science*, 58(2), 207–33.
Willaschek, M. (1992) *Praktische Vernunft: Handlungstheorie und Moralbegründung bei Kant*. Stuttgart: J.B. Metzler.
Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
Zhang, J. (1997) 'The nature of external representations in problem solving'. *Cognitive Science*, 21, 179–217.

Cite this article: Rachavelpula, S. (2026). Modelling Morality: A Kantian Account of Moral Examples. *Kantian Review* **31**, 71–92. https://doi.org/10.1017/S1369415426101198

https://doi.org/10.1017/S1369415426101198 Published online by Cambridge University Press