

Monica Bucciarelli, Philippe N. Johnson-Laird

Deontics: Meaning, Reasoning, and Emotion

(doi: 10.1436/92959)

Materiali per una storia della cultura giuridica (ISSN 1120-9607)

Fascicolo 1, giugno 2019

Ente di afferenza:

Universitgli studi di Genova (unige)

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

DEONTICS: MEANING, REASONING, AND EMOTION

by Monica Bucciarelli and Philip N. Johnson-Laird

This article reports psychological experiments that corroborate the following account. Deontic reasoning relies on mental models of possibilities in a deontic context. Because these models represent what is permissible rather than impermissible, individuals commit predictable fallacies in reasoning from certain sorts of deontic premise. Contrary to a tradition going back to Hume, humans reason in order to make moral judgments. Their inferences can be rapid, intuitive, and based on a single model, but they can also be slow, deliberative, and based on alternative models, as when they resolve a dilemma. Humans have an innate system of basic emotions, which is inherited from our evolutionary ancestors. These emotions are elicited by primitive cognitions that are too crude to distinguish between causes and enabling conditions. The distinction calls for a deliberative inference so subtle that some learned jurists have not realized that the two concepts differ in meaning. Unlike factual propositions, moral propositions have a striking relation with emotions. People love those moral propositions that they believe, and hate those that they disbelieve. The effect can be elicited from the mere substitution of the word *ought* for *is* in an assertion. In sum, a comprehensive theory of deontics must account for meaning, reasoning, and emotion.

Keywords: Beliefs, Deontics, Emotions, Mental Models, Moral Judgments.

Foreword

Deontic propositions are central to laws, rules, conventions, morality, etiquette, and so on and on. They concern what is permissible, what is obligatory, and their denials. A crucial difference between deontic and factual propositions is that a violation of a factual proposition shows that the proposition is false, whereas a violation of a deontic obligation has no bearing on its truth or falsity – it shows that an individual is in breach of the obligation, not that it

Monica Bucciarelli. Dipartimento di Psicologia and Center for Logic, Language and Cognition, Università di Torino, Via Po 14, 10123, Torino, monica.bucciarelli@unito.it

Philip N. Johnson-Laird, Stuart Professor of Psychology, Emeritus, Princeton University; Visiting scholar, Department of Psychology, New York University, phil@princeton.edu

is false¹. The present article aims to elucidate deontics and morality, and to explain the roles of reasoning and emotions in them. It has three parts. First, it outlines a theory of the meaning of deontic assertions. It explains how they are based on an underlying system of possibilities, and it describes their role in reasoning. Second, it considers the special class of deontic assertions that concern morality. It argues that they have no simple defining characteristics that separate them from other sorts of deontic assertions. Yet, contrary to some accounts, notably Hume's and his followers', people do reason in order to make moral judgments. Third, it considers the relations between morality and emotions – Hume's criterion for moral judgments. Experiments show that people love those moral propositions that they believe, and hate those that they disbelieve. No such relation exists for factual propositions. But, it does not follow that moral judgments rely on emotions. As we will show, reasoning and emotions are two independent systems, and deontic evaluations, including those about morality, depend on reasoning.

1. *Meaning, models, and deontic reasoning*

Deontic assertions can be expressed in terms of possibilities and necessities, e.g., *It is possible for you to leave work early today, It is necessary for you to start work early tomorrow*. Here, the use of “possible” and “necessary” shows that these modal terms embrace deontic concepts: deontic logics are accordingly a branch of modal logics². But, modal logics base the meanings of their concepts on “possible worlds”, where each possible world determines whether any assertion is true or false. Possible worlds are therefore too big to fit inside anyone's head³. A more plausible alternative is that the meanings of modal terms in daily life have the same basis as everyday probabilities. In any situation that individuals can think of as having a small number of exhaustive and mutually exclusive alternatives, a *possibility* refers to a subset of these alternatives, and a *necessity* refers to all of them⁴. Hence, deontic meanings concern what is possible or necessary according to the relevant laws, rules, social

¹ R. Wertheimer, *The Significance of Sense: Meaning, Modality and Morality*. Ithaca (NY), Cornell University Press, 1972.

² See, e.g., R. Girle, *Modal Logics and Philosophy*, London, Routledge, 2009.

³ B. Hall-Pardee, *Semantics – Mathematics or Psychology?*, in R. Bäuerle, U. Egli, A. von Stechow, eds., *Semantics from Different Points of View*, Berlin, Springer-Verlag, 1979.

⁴ M. Ragni, P.N. Johnson-Laird, *Possibilities: a Theory of Naive Modal Reasoning*, 2019, under submission.

conventions, and so on. An immediate consequence of this semantics is that, as in deontic logics, the two concepts of *permissible* and *obligatory* are interdefinable using negation:

- (1) A is permissible if, and only if, not-A is not obligatory.
A is obligatory if, and only if, not-A is not permissible.

There are therefore only four absolute deontic assertions:

- (2) B is permissible: it is a deontic possibility.
Not-B is permissible: it is not a deontic necessity.
B is obligatory: it is a deontic necessity.
B is impermissible: it is not a deontic possibility.

Deontic assertions range from absolute moral principles, such as:

- (3) Thou shalt not kill

to relational claims, such as:

- (4) The judge's approval permits the prisoner to be paroled.

Relational claims have interpretations in common with one of four sorts of deontic relation:

- (5) A permits B.
A permits not-B.
A obligates B.
A prohibits B.

where *A* and *B* stand for actions or inactions. Normally, *A* occurs before *B*, but there are cases in which permissions are granted retrospectively. The four deontic relations in (5) are basic, because they underlie any deontic discourse, and because they exhaust the set of possible binary deontic relations. But, they can be expressed in many different ways, e.g.:

- (6) If the judge approves, then the prisoner can be paroled.
The prisoner is allowed parole if it has the judge's approval.
With the judge's approval, the prisoner is permitted parole.

The theory of mental models – the “model” theory for short – postulates that individuals reason by constructing mental models of

the world. A mental model is a representation that insofar as possible has the same structure as the situation that it represents, which may be static or kinematic⁵. Real mental models represent the world, but, for simplicity, diagrams that denotes mental models use words and phrases. A deontic claim such as:

(7) If you have children then you must earn a living.

has mental models of the conjunction of deontic possibilities to which it refers. Mental models of a conditional, however, do not represent explicitly the possibilities in which the *if*-clause does not hold:

(8) have children earn a living

. . .

The first row in this diagram represents the possibility in which you have children, in which case you earn a living – it is the only possibility given that you have children. The ellipsis in the second row denotes the possibilities in which you do not have children, but it has no explicit content. Mental models accordingly abide by a principle of deontic mental models: they represent what is permissible but not normally what is impermissible. An analogous principle of truth governs mental models of epistemic possibilities. In simple situations, individuals are able to flesh out mental models into *fully explicit* models. The fully explicit models of the possibilities to which the conditional (7) refers are as follows, where the symbol, “ \neg ”, denotes negation:

(9) have children earn a living
 \neg have children earn a living
 \neg have children \neg earn a living

Fully explicit models also enable reasoners to consider the case in which the deontic relation is violated:

(10) have children \neg earn a living

⁵ P.N. Johnson-Laird, *Mental Models*, Cambridge, Cambridge University Press, 1983; P.N. Johnson-Laird, *How We Reason*, Oxford, Oxford University Press, 2006; P.N. Johnson-Laird, R.M.J. Byrne, *Deduction*, Hillsdale (NY), Erlbaum, 1991.

The theory calls for models of many different sorts: «what underlies [...] meaning is the ability to envisage states of affairs that may or may not correspond to reality, that is, the ability to construct mental models of possible, hypothetical, and imaginary situations»⁶. They can represent different sorts of discourse – factual, hypothetical, or fictional. And they can represent real possibilities, counterfactual possibilities, and deontic possibilities. There is accordingly a system of symbols attached to models that signify their status, just as negation has to be represented in models by a symbol that has an associated semantics⁷. As far as possible, the present article avoids the use of symbols on models. The models with which it deals refer to deontic possibilities, and in general what is deontically possible is also epistemically possible, i.e., an event or situation that can occur in the light of the speaker’s knowledge⁸. Hence, a full paraphrase of (7) makes clear these relations:

(11) It is possible that you have children and if you do then the one and only deontic possibility is that you earn a living, which it is possible that you do.

If there is only a single deontic possibility, it is an obligation.

Knowledge according to the model theory can modulate the interpretation of assertions. Assertion (7) is a weak obligation, because *A* obligates *B*, but *B* can occur in the absence of *A*. A strong obligation occurs in this assertion:

(12) If you have children then you are obligated to look after them.

It has only two models, because the pronoun “them” refers to your children, and so you cannot look after your children unless you have children:

(13) have children look after them
 ¬ have children

In the second case, the pronoun “them” has nothing to refer to, and so the clause is vacuous. Table 1 presents the sets of deontic pos-

⁶ P.N. Johnson-Laird, *Mental Models*, cit., p. 60.

⁷ P.N. Johnson-Laird, R.M.J. Byrne, *Deduction*, cit., pp. 68-69.

⁸ M. Ragni, P.N. Johnson-Laird. *Possibilities*, cit.

sibilities for the four basic relations in both their weak and strong interpretations.

Table 1. The sets of deontic possibilities for the four deontic relations in both weak interpretations (on the left) and strong interpretations (on the right), where ‘+’ indicates a viable interpretation⁹.

The four sorts of assertions	The sets of deontic possibilities to which the assertions refer							
	A B	A \neg B	\neg A B	\neg A \neg B	A B	A \neg B	\neg A B	\neg A \neg B
A permits B	+	+						
A permits not-B	+		+					
A obligates B					+	+		
A prohibits B							+	+

When individuals understand deontic assertions, they tend to use mental models rather than the fully explicit models in Table 1, because mental models impose less of a load on working memory. The model theory accordingly distinguishes between intuitive inferences based on single mental models, which are considered one at a time, and deliberative inferences based on fully explicit models, which allow individuals to consider alternative models. Such a distinction between two systems of reasoning was proposed first by the late Peter Wason¹⁰, and Wason’s students are among those who have developed such “dual process” theories of reasoning¹¹, but others have also formulated them¹². The computer program, mSentential, implementing the model theory is perhaps the only program to simulate the two systems of reasoning¹³.

The principle of deontic mental models – an analog to the principle of truth in the factual domain – captures the distinction between mental models and fully explicit models. Any given relation in

⁹ After M. Bucciarelli, P.N. Johnson-Laird, *Naïve Deontics: A Theory of Meaning, Representation, and Reasoning*, in «Cognitive Psychology», 50, 2, 2005, pp. 159-193.

¹⁰ P.C. Wason, P.N. Johnson-Laird, *A Conflict Between Selecting and Evaluating Information in an Inferential Task*, in «British Journal of Psychology», 61, 1970, pp. 509-515.

¹¹ E.g. J.St.B.T. Evans, *Thinking Twice: Two Minds in One Brain*, Oxford, Oxford University Press, 2010; P.N. Johnson-Laird, *Mental Models*, cit., ch. 6.

¹² D. Kahneman, *Thinking: Fast and Slow*, London Penguin, 2011.

¹³ See <http://mentalmodels.princeton.edu/models/>.

Table 1 can be expressed in several ways. For example, *A prohibits B* refers to the same possibilities as *Not-A permits B*. The two assertions have the same fully explicit models, but their mental models differ. The mental models of *A prohibits B* are as follows:

- | | |
|-------------------------------------|-------------------------|
| (14) Permissible
A ¬ B
... | Impermissible
A B |
|-------------------------------------|-------------------------|

In contrast, the mental models of *Not-A permits B* are as follows:

- (15) Permissible
 ¬ A B
 ¬ A
 ...

As (15) makes evident, there are no mental models of what is impermissible for assertions about permissions. Their recovery calls for deliberative inferences. In other words, only prohibitions have mental models of what is impermissible, because prohibitions are the only sort of deontic assertion that makes them salient.

Bucciarelli and Johnson-Laird¹⁴ tested the model theory's predictions about the fully explicit models of deontic assertions. The participants in an experiment had to list what was permissible and what was impermissible for assertions expressing two versions of each of the four basic deontic relations in Table 1. For example, here are the sentences expressing two versions of the same relation:

- (16) Tax-payers who support charities are permitted to claim a rebate on their taxes.

and:

- (17) Tax-payers who do not support charities are prohibited from claiming a rebate on their taxes.

The results corroborated the theory. The participants tended to list the fully explicit possibilities shown in Table 1, and the lists were the same for the two ways of expressing each relation. But, they tended to start their lists with the possibility represented in the

¹⁴ M. Bucciarelli, P.N. Johnson-Laird, *Naïve Deontics*, cit.

mental model of an assertion. For (16), the list started with the possibility corresponding to *A and B*, whereas for (17) it tended to start with the possibility corresponding to *not-A and the impermissibility of B*. The participants, however, were biased towards weak interpretations of assertions referring to permissions, that is, they listed all four cases as permissible.

Reasoning relies on working memory – a short-term memory that holds information in mind for this and other processes, and so the model theory predicts that deductions should be based on mental models rather than fully explicit models. Reasoners should therefore tend to draw those conclusions that hold in their mental models of the premises. Consider this pair of premises:

- (18) Having children obligates you to look after them.
(A obligates B)

To look after children prohibits you from leaving them unattended.
(B prohibits C)

The conjunction of the two premises yields the following mental models:

- (19) Have children look after them \neg leave unattended
 . . .

Granted that reasoners aim to draw conclusions that maintain the information in the premises, and to formulate a conclusion that is not explicit in them¹⁵, they should tend to draw the conclusion:

- (20) Therefore, having children prohibits you from leaving them unattended.
(A prohibits C)

Bucciarelli and Johnson-Laird¹⁶ corroborated the model theory's prediction for 16 different sorts of pairs of premises. The participants' spontaneous conclusions tended to be those that the mental models of the premises predicted.

¹⁵ P.N. Johnson-Laird, R.M.J. Byrne, *Deduction*, cit.

¹⁶ M. Bucciarelli, P.N. Johnson-Laird, *Naïve Deontics*, cit.

The most striking corroboration of the model theory is the occurrence of systematic fallacies that the theory predicts. They can be so compelling that they have the character of cognitive illusions. For example, consider the following inferential problem:

(21) You are permitted to carry out only one of the following actions:

Action 1: take the apple or the orange, or both

Action 2: take the pear or the orange, or both

Are you permitted to take the orange?

Most people say, "Yes". The mental models of action 1 represent the three permissible actions. You can take one or other of the fruits or both of them:

(22) apple

orange

apple

orange

These models support the conclusion that it is permissible to take the orange. The mental models of action 2 support the same conclusion:

(23) pear

orange

pear

orange

So, reasoners should respond:

(24) Yes, I'm permitted to take the orange.

The response is an illusion. The problem gives you permission to carry out only one action, but if you take the orange then you are carrying out both actions. So, when one action is permissible the other action is impermissible. The fully explicit models of the premises are therefore:

(25) apple

\neg pear

\neg orange

\neg apple

pear

\neg orange

They show that you can take the apple or the pear, but you cannot take the orange. A simple control inference for the illusion has the same premises, but poses a different question:

(26) Are you permitted to take the pear?

The mental models now yield the correct inference:

(27) Yes, I'm permitted to take the pear.

An experiment examined deontic problems with illusory “yes” responses and illusory “no” responses, and their respective controls. The participants made just 7% correct responses to the illusions, but 90% correct responses to the control inferences¹⁷. Because the control inferences have the same premises as the illusory inferences, and the questions differ in just one word, such as “pear” in place of “orange”, the results cannot be due to the artificiality of the problems.

The model theory predicts an antidote to illusory inferences. When individuals deal with the deontic verb, *prohibits*, they should think first of the states of affairs that are impermissible. Consider this problem:

(28) You are prohibited from carrying out more than one of the following actions:

Action 1: take the apple or the orange, or both

Action 2: take the pear or the orange, or both

Are you permitted to take the orange?

Individuals should tend to construct models of what is impermissible:

(29) apple pear
 orange

So, you can't take both the apple and pear, and you can't take the orange. An experiment showed that although illusory problems remained harder than controls, the participants did much better with illusory problems based on prohibitions than with those based on permissions¹⁸.

Overall, the experimental results support the model theory's predictions, and present three challenges to alternative theories. They need to account for the meanings of modal terms, for the salience of some deontic possibilities over others, and for the occurrence of illusory inferences. This last problem is severe for theories based on

¹⁷ *Ibidem.*

¹⁸ *Ibidem.*

deontic logic¹⁹. Rules of inference that yield only valid conclusions cannot explain the illusions; rules of inference that can explain the illusions cannot be consistent with those that yield valid conclusions.

2. Reasoning about morality

Deontic propositions often concern morality. When individuals make a moral evaluation, a sensible question is to what extent, if any, they rely on reasoning. The Scottish Empiricist Hume²⁰ denied any role for reasoning, which he took to be incapable of initiating action. He argued that moral judgments depend on emotions. In contrast, Kant²¹, the founder of Rationalism, based morality on a categorical imperative – that one should act according to a principle that at the same time you intend to be a universal law. He sought a priori moral principles that apply the categorical imperative to all humanity regardless of culture. These two views have counterparts in current psychological theories. According to the “socio-intuitionist” theory²², reasoning plays no role in making moral judgments. They rely instead on emotional reactions of approval or disapproval. Insofar as reasoning enters the process it occurs after the judgment, and it is effortful, conscious, and serves to persuade others²³. Followers of the Rationalist tradition postulate instead that moral judgments are unconscious intuitions from an innate moral grammar, which has been fine-tuned by cultural experience²⁴. This theory rests on an analogy between morality and Chomsky’s²⁵ conception of a universal grammar for natural language that is fine-tuned by a speaker’s native tongue. Such a grammar is consistent: it never treats a given sequence of words as both grammatical and ungrammatical. Likewise, a moral grammar should never evaluate an action as both

¹⁹ L.J. Rips, *The Psychology of Proof*, Cambridge (MA), MIT Press, 1994, p. 322, et seq.

²⁰ D. Hume, (1978), *A Treatise of Human Nature*, Oxford, Oxford University Press, 1978 (originally published 1739), Book III.

²¹ I. Kant, *Foundations of the Metaphysics of Morals, and What Is Enlightenment?*, trans. by L. W. Beck, Indianapolis, Bobbs-Merrill, 1959 (originally published 1785).

²² J. Haidt, *The Emotional Dog and its Rational Tail*, in «Psychological Review», 108, 2001, pp. 814-834; J. Haidt, *The New Synthesis in Moral Psychology*, in «Science», 316, 2007, pp. 998-1002.

²³ J. Haidt, J. Graham, *When Morality Opposes Justice: Conservatives Have Moral Intuitions That Liberals May Not Recognize*, in «Social Justice Research», 20, 2007, pp. 98-116.

²⁴ M.D. Hauser, *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*, New York, NY, Harper Collins, 2006; J. Mikhail, *Universal Moral Grammar: Theory, Evidence and the Future*, in «Trends in Cognitive Sciences», 114, 2007, pp. 143-152.

²⁵ N. Chomsky, *Knowledge of Language: Its Nature, Origin, and Use*, Westport (CT), Praeger, 1986.

permissible and impermissible. And if the grammar is complete, it should always decide whether any action in the domain of morality is right or wrong. Like the socio-intuitionist theory, this theory also downplays the role of reasoning in reaching moral decisions. Both accounts yield rapid intuitions that are more akin to perceptions than cognitions.

Not every current theory eschews reasoning²⁶. Likewise, the extension of the model theory to deontics postulates that moral judgments depend on reasoning – both the system yielding intuitions and the system yielding deliberations²⁷. This section of the article aims to make progress towards deciding which of these approaches is correct. It outlines a theory of emotions that relates emotions to models, it recounts the difficulty of identifying what counts as a moral proposition, it describes how the model theory embodies moral reasoning, and finally it reports experiments designed to decide among the three sorts of theory.

Emotions are communications both within the brain and among individuals²⁸. Darwin²⁹ was the earliest advocate of this claim, and he realized that from an evolutionary standpoint basic emotions are at least as old as social mammals. They are innate in human beings and have their own distinctive signals in the brain and in universal facial expressions³⁰. Their signals in all organisms prepare them for various courses of action relating to the ontogeny of the species. The paradigm examples of basic emotions in humans are *happiness*, *sadness*, *anger*, *anxiety*, and *disgust*. Humans also experience them in ways that are indissolubly bound to propositional contents. These

²⁶ E.g. J. Baron, *Thinking and Deciding*, New York, Cambridge University Press, 2008, ch. 16; J.D. Greene, R.B. Sommerville, L.E. Nystrom, J.M. Darley, J.D. Cohen, *An fMRI Investigation of Emotional Engagement in Moral Judgment*, in «Science», 293, 2001, pp. 2105-2108; J.M. Paxton, L. Ungar, J.D. Greene, *Reflection and Reasoning in Moral Judgment*, in «Cognitive Science», 36, 2011, pp. 163-177.

²⁷ M. Bucciarelli, *Moral Dilemmas in Females: Children Are More Utilitarian Than Adults*, «Frontiers in Psychology», 6, 2015, p. 1345, <http://doi.org/10.3389/fpsyg.2015.01345>; M. Bucciarelli, M. Daniele, *Reasoning in Moral Conflicts*, «Thinking & Reasoning», 21, 2014, pp. 265-294; M. Bucciarelli, S. Khehlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, in «Judgment and Decision Making», 3, 2, 2008, pp. 121-139.

²⁸ K.J. Oatley, P.N. Johnson-Laird, *Towards a Cognitive Theory of Emotions*, in «Cognition & Emotion», 1, 1987, pp. 29-50; K.J. Oatley, P.N. Johnson-Laird, *The Communicative Theory of Emotions: Empirical Tests, Mental Models, and Implications for Social Interaction*, in L.L. Martin, A. Tesser, eds., *Striving and Feeling: Interactions Among Goals, Affect, and Self-regulation*, Mahwah (NY), Erlbaum, 1996, pp. 363-393.

²⁹ C. Darwin, *The Expression of the Emotions in Man and the Animals*, Chicago, University of Chicago Press, 1965 (originally published in 1872).

³⁰ D. Keltner, P. Ekman, G.C. Gonzaga, J. Beer, *Facial Expression of Emotion*, in R.J. Davidson, K.R. Scherer, H.H. Goldsmith, eds., *Handbook of Affective Sciences*, New York, Oxford University Press, 2003, pp. 415-432.

“complex” emotions concern mental models of the self and others³¹. Unlike certain basic emotions, individuals therefore cannot experience complex emotions without being aware of the evaluations that precipitated them. They include such emotions as *remorse*, *jealousy*, and *pride*. Remorse, for instance, is sadness about an action or inaction that harms another, and individuals have this emotion because they judge themselves to have violated the moral code embodied in their idealized models of themselves. Other complex emotions have analogous ties with propositional contents. And, as the next section shows, this tie between emotions and contents is crucial for morality.

Moral propositions are deontic, but a major puzzle is what singles them out from other deontic propositions. No simple non-circular way seems to exist to do so. Doubtless morality has an innate basis, but what is right and what is wrong differs from one culture to another, and even from one individual to another. There are even differences in what counts as a moral issue. Is smoking a moral issue? Is over-eating a moral issue? Is smacking a child a moral issue? For some people, they are all morally wrong. For others, they are not even moral issues – like bird watching, they are neither right nor wrong. Likewise, what is a crime in one culture can be quite moral in another, e.g., bigamy, charging interest on a loan, and killing unbelievers. Someone may yet frame a simple criterion for identifying all and only moral propositions. But, efforts to do so in philosophy and psychology all seem to fail, e.g.:

- moral propositions concern welfare, justice, and rights;
- they concern violence, dishonesty, and cruelty;
- they are immutable principles that apply to everyone;
- they concern violations that merit punishment;
- they concern the rightness or wrongness of actions that knowingly harm others.

The authors of these accounts are identified elsewhere³². It is remarkable how often their efforts are inapplicable to morally good actions, and how often they presuppose what they are trying to analyze. The concepts of justice, rights, punishment, and rightness and wrongness, are all moral concepts. So far, there exists neither a way to demarcate moral propositions nor, despite claims in the media, evidence for a brain mechanism dedicated only to moral judgments. A plausible conclusion is that morals depend on beliefs, which indi-

³¹ P.N. Johnson-Laird, K. Oatley, *Cognitive and Social Construction in Emotions*. In M. Lewis, J.M. Haviland-Jones, eds., *Handbook of Emotions*, New York, Guilford Press, 2000², pp. 458-475.

³² M. Bucciarelli, S. Khemlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, cit.

viduals acquire from their parents and peers, and these beliefs determine which propositions are about morality³³.

The model theory of morality rests on three main principles³⁴. First, it postulates that the foundations of moral judgments are neither complete nor consistent (*the principle of moral inconsistency*). Beliefs are the basis of morals, and beliefs are liable to be inconsistent. To establish their consistency is computationally intractable³⁵. Beliefs depend on “atomic” propositions, such as: *you have children*, and *you earn a living*, which can each be true or false. A set of beliefs based on three atomic propositions (*A*, *B*, and *C*) of the sort: *If A then B*, *If B then C*, and *If C then not A*, is inconsistent: the three beliefs cannot all be true. And if a set of beliefs depends on, say, 100 atomic propositions, then the task can call for checking that each of 2^{100} distinct combinations of them is consistent. That number is vast (1 followed by 30 digits), and even if one could check one combination per second, it would take longer than the universe has existed to check their consistency. So, beliefs are bound to risk inconsistency. A corollary is the principle of moral inconsistency. It predicts inconsistencies should occur in moral evaluations and in deontic systems. That is why moral dilemmas exist. For instance, criminal lawyers are obligated to keep confidential their client’s disclosures and yet to be candid in court. So, they are in a real dilemma if they know that a client has committed perjury³⁶.

Second, the model theory postulates that deontic evaluations, including those concerning morality, depend on reasoning (*the principle of deontic reasoning*). And reasoning depends on intuitions or deliberations (see the previous section), and its processes are the same as those for factual reasoning. In other words, no special sort of reasoning exists for dealing with deontic matters – deontic reasoning is normal reasoning but about deontic topics³⁷.

Third, the model theory postulates that emotions and reasoning are two independent systems operating in parallel (*principle of independent systems*). Granted their evolutionary origins, emotions ante-

³³ M. Bucciarelli, P.N. Johnson-Laird, *Naïve Deontics*, cit.

³⁴ M. Bucciarelli, S. Khemlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, cit.

³⁵ S.A. Cook, *The Complexity of Theorem Proving Procedures*, in «Proceedings of the Third Annual Association of Computing Machinery Symposium on the Theory of Computing», 3, 1971, pp. 151-58.

³⁶ T. McConnell, *Moral Dilemmas*, 2014, in E.N. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/moral-dilemmas/>.

³⁷ Cf. F. Cushman, L. Young, *Patterns of Moral Judgment Derive from Nonmoral Psychological Representations*, in «Cognitive Science», 35, 2011, pp. 1052-1075; K. Uttich, T. Lombrozo, *Norms Inform Mental State Ascriptions: A Rational Explanation for the Side-Effect Effect*, in «Cognition», 116, 2010, pp. 87-100.

date deontic evaluations by several hundred million years. An emotional reaction may occur to a situation that has no deontic status, such as a landscape or an earthquake. Conversely, the theft of a paper napkin, can elicit a deontic evaluation but little or no emotional reaction. Of course the two systems interact, and complex emotions, such as remorse, bind together a basic emotion and a propositional content concerning morality.

Experiments support the model theory. Given a scenario to evaluate, individuals can make a rapid deontic evaluation and then reason consciously. Such a sequence is consistent with an intuition preceding deliberative reasoning. But, another sequence is a chain of conscious reasoning culminating in an evaluation. Still another sequence is a snap moral evaluation followed at once with a claim explaining its reasons. This sort of sequence is ambiguous between the two previous cases. One study used a series of scenarios, which each described a single outcome, either moral or immoral³⁸. Two agents played distinct causal roles in each scenario: the action of one agent enabled the action of the other to cause the outcome. Previous studies had shown that naive individuals distinguished between the two sorts of agents³⁹. Yet, the distinction between causers and enablers is so subtle that the law in the English-speaking world takes the view that no principled difference exists between the two⁴⁰ – the difference between them is, in Mill's term, capricious⁴¹. A typical scenario from the experiment was:

(30) Barnett owned a gun store. He sold guns to everyone without checking IDs or whether the buyer had a criminal record. Martin came into the store intending to buy a weapon, and left with a handgun. He went home and fired it repeatedly. Later, his wife died from her wounds.

The participants judged which of the two agents was more blameworthy for immoral outcomes, and which of the two was more praiseworthy for moral outcomes. But, one of the two groups of

³⁸ M. Bucciarelli, S. Khemlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, cit.

³⁹ C. Frosch, P.N. Johnson-Laird, M. Cowley, *It's Not My Fault, Your Honor; I'm Only the Enabler*, in D.S. McNamara, J.G. Trafton, eds., *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, Nashville, Tennessee, USA, 2007, p. 1755.

⁴⁰ H.L.A. Hart, A.M. Honoré, *Causation in the Law*, Oxford, Clarendon Press, 1985².

⁴¹ J.S. Mill, *A System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, Toronto, University of Toronto Press, Routledge & Kegan Paul, 1973 (originally published in 1843). For an account of their difference in meaning, and its legal ramifications, see P.N. Johnson-Laird, *Causation, Mental Models, and the Law*, in «Brooklyn Law Review», 65, 1999, pp. 67-103.

participants had to think aloud as they made their decision. The protocols of people thinking aloud are a fairly reliable guide to their sequences of thought⁴². The participants in both groups tended to chose the causer rather than the enabler as more praiseworthy for good outcomes and more blameworthy for bad outcomes (83% in both groups). Nearly all the protocols from the participants thinking aloud fell into one of the three categories described above. And most of them showed that the participants tended to reason in order to make their judgments. A typical protocol for the scenario above included the following thoughts:

- (31) [...] ultimately Martin is the one who made the decision to [...] commit the crime, but Barnett is the one who supplied the guns and, by law, Barnett would also be at the same level of blame. But morally I feel that Barnett should [...] is definitely less blameworthy than Martin because he sold [...] he sells the guns and ultimately it's the decision of the consumer or whoever buys it how to use it.

To make sense of the scenario, the participant had to infer that Barnett sold a handgun to Barnett, and that Barnett shot his wife with the handgun. Neither proposition is asserted in the text of the scenario (30). Hence, this participant, like the others, is reasoning. Likewise, this participant deliberates in order to move from the view that the two agents are equally blameworthy in the law to the view that the one who enabled the outcome is less blameworthy than the one who caused it. The sequence of thoughts anticipates the moral judgment, and it is plain that it yields the final moral evaluation.

A study of the evaluation of moral scenarios that contained conflicts corroborated the role of deliberative reasoning⁴³. The participants' task was to decide whether the actions in a scenario were right or wrong. It took them longer to evaluate scenarios describing morally ambiguous actions than those describing actions that were clearly morally right or morally wrong. A further study examined what participants had to say in thinking aloud as they made their decisions. The participants tended to make intuitive evaluations of unambiguous scenarios, whereas they tended to make reasoned evaluations of conflicting scenarios. The experiment corroborated

⁴² K.A. Ericsson, H.A. Simon, *Verbal Reports as Data*, in «Psychological Review», 87, 1980, pp. 215-251.

⁴³ M. Bucciarelli, M. Daniele, *Reasoning in Moral Conflict*, cit.

the principle that people reason to make moral evaluations, and it demonstrated the role of both intuitive and deliberative reasoning.

Individuals are able to modify moral dilemmas to make them difficult, if not impossible, to resolve⁴⁴. The participants in an experiment judged that in certain scenarios an agent acted in a way that was morally wrong, e.g., the driver of a motorboat pushed a passenger overboard in order to save some swimmers. They were then able to edit these dilemmas to turn them into irresolvable dilemmas, e.g., they changed the swimmers into the driver's children. A follow-up study showed that the construction of such dilemmas depends, not on modifying the emotions that a dilemma elicits, but on modifying propositions relating to moral principles⁴⁵. The participants judged various scenarios as morally unambiguous, e.g.:

- (32) A single woman took care of twenty stray dogs that would otherwise have been put in a kennel and left to die, treating them with great care and love.

They then had to change the scenario into one that was hard to judge as morally right or morally wrong. They were more likely to change propositions pertinent to moral principles than to change propositions evoking emotional reactions. For instance, a participant modified the preceding scenario (32) in the following way:

- (33) A single woman took care of twenty stray dogs that otherwise would have been left to die, training them to be aggressive.

This change violates a norm: it is wrong to train pets to be aggressive. A change invoking an emotion would have been, for example:

- (34) A single woman took care of twenty stray dogs that otherwise would have been left to die, in order to feel superior to her friends.

Contrary to the hypothesis that emotions mediate moral judgments, the participants had a massive bias towards introducing propositions pertinent to moral norms rather than to emotional consequences. The results supported the principle that emotional and deontic systems are independent.

⁴⁴ M. Bucciarelli, S. Khemlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, cit.

⁴⁵ M. Bucciarelli, M. Daniele, *Reasoning in Moral Conflict*, cit.

When individuals react to moral scenarios, they agree with one another that some scenarios elicit first an emotional reaction, other scenarios elicit first a moral evaluation, and still others scenarios have no bias either way in what they elicit⁴⁶. Emotions came first for positive scenarios about love, kindness, and friendship, and for negative scenarios about violence and other horrifying topics. Moral evaluations came first for positive scenarios about helping disabled individuals, and for negative scenarios about bribery, perjury, and other crimes without violence. The positive scenarios for which there was no bias either way were about cooperation or care, and the negative scenarios were about crimes against property or sexual topics. A follow up showed corresponding latencies in the participants' answers to questions, e.g., they were faster to answer questions about emotions for the scenarios for which emotions came first in the previous study. These results also corroborated the principle of independent systems for emotions and deontics.

What does all the evidence imply about the alternatives to the model theory? The socio-intuitionist theory postulates that moral evaluations are based on rapid emotional intuitions. The experimental results are incompatible with this theory. They show that people do reason, and even deliberate, in order to make moral judgments, that when they are challenged to create a moral dilemma out of an unambiguous scenario, they change propositions about moral precepts rather than propositions about emotional reactions, and that they react to some moral scenarios with an emotion first, but to others with a moral evaluation first. Perhaps the biggest problem for the notion that emotions yield moral evaluations is that the system of emotions in itself, which is common to other social mammals, lacks the propositional equipment to make intricate moral judgments. It cannot tell the difference between causes and enabling conditions.

The theory of moral grammar does no better. If moral intuitions rest on a grammar, then insoluble dilemmas should be impossible. But, individuals can change a scenario into a moral dilemma. That result is inconceivable if moral evaluations are intuitions based on a grammar. It should yield only morally right or wrong evaluations, or, if it is incomplete, no evaluation at all. Likewise, the demonstrable role of time-consuming deliberative reasoning in reaching a moral decision is contrary to grammatical intuitions. No theory of moral grammar has outlined a substantial set of its rules, and so it is impossible to assess whether the theory can accommodate the vast differences

⁴⁶ M. Bucciarelli, S. Khemlani, P.N. Johnson-Laird, *The Psychology of Moral Reasoning*, cit.

from one person to another about which issues are moral. Likewise, the theory seems to say nothing about the relations between morality and emotions. In contrast, the model theory postulates two independent systems that can interact with one another. In their experiments, Bucciarelli and Daniele⁴⁷ observed a correlation between the strength of participants' moral and emotional reactions. So, perhaps Hume and Haidt were right about an intimate relation between morality and emotions. The final part of the article takes up this possibility.

3. *Emotions and Morality*

A common observation is that people are emotionally attached to certain of their beliefs. This attachment was referred to earlier as a “complex” emotion, which binds together a basic innate emotion to propositional content. These attachments are obvious in the case of religious and political beliefs: unconscious inferences lead to a bond between emotions and beliefs. The beliefs are about things unseen, but they are vital, and the emotional tie to them can render them immune to reasoned argument. Individuals give up mundane beliefs when facts clash with them, but matters are quite different with religious beliefs:

Why are these untestable beliefs so powerful that those who hold them would sooner die than abandon them? The answer [...] depends on an unconscious transition to an emotion. The beliefs concern how we should live, the nature of death, and survival beyond it. These beliefs have an extraordinary capacity to invoke a basic emotion of attachment. We become attached to God the father and mother church. And basic emotions [...] are created by simple evaluations leading to unconscious transitions. Any challenge to these primeval beliefs is a deep threat, and it will be resisted⁴⁸.

The connections between emotions and beliefs have been a matter for speculation⁴⁹. And studies have examined various relations⁵⁰.

⁴⁷ M. Bucciarelli, M. Daniele, *Reasoning in Moral Conflict*, cit.

⁴⁸ P.N. Johnson-Laird, *How We Reason*, cit., p. 334.

⁴⁹ E.g. G.L. Clore, K. Gasper, *Feeling in Believing: Some Affective Influences on Belief*, in N.H. Frijda, A.S.R. Manstead, S. Bem, eds., *Emotions and Beliefs: How Feelings Influence Thoughts*, Cambridge, Cambridge University Press, 2000, pp. 10-44; K. Fielder, H. Bless, *The Formation of Beliefs at the Interface of Affective and Cognitive Processes*, in N.H. Frijda, A.S.R. Manstead, S. Bem, eds., *Emotions and Beliefs: How Feelings Influence Thoughts*, Cambridge, Cambridge University Press, 2000, pp. 133-170; N. H. Frijda, A.S.R. Manstead, S. Bem, *Emotions and Beliefs: How Feelings Influence Thoughts*, Cambridge, Cambridge University Press, 2000, pp. 1-9; C.I. Hovland, I.L. Janis, H. H. Kelley, *Communication and Persuasion: Psychological Studies on Opinion Change*, New Haven, Yale University Press, 1953; J. Mercer, *Emotional Beliefs*, in «International Organization», 64, 1, 2010, pp. 1-31.

⁵⁰ E.g. J. Moll, R. de Oliveira-Souza, I.E. Bramati, J. Grafman, *Functional Networks in Emotional Moral and Non-Moral social Judgments*, in «Neuroimage», 16, 2002, pp. 696-703;

In experiments using brain imaging, for example, the participants considered information consistent or inconsistent with their political beliefs. When they considered consistent information, brain regions that mediate emotions were highly active⁵¹. Yet, we can have a strong belief in a proposition without having any emotional attachment to it. We may be certain, for instance, that the time is 9.30am and be quite indifferent to the fact. Such a relation raises the question of what underlies the “strength” of a belief. In the present authors’ view, it is the likelihood of the belief, that is, its subjective probability. This idea goes back to Ramsey⁵², de Finetti⁵³, and other defenders of subjective probability – a viewpoint sometimes known as Bayesianism. The mental mechanism underlying subjective probabilities, especially those for unique propositions, such as that Trump will be impeached, is based on the proportion of models of evidence, which the proposition elicits, in which the event occurs⁵⁴.

Our moral beliefs seem more akin to religious beliefs than to factual beliefs: after all, many moral precepts derive from creeds. Hence, if we believe a moral proposition we should like it, and if we don’t believe a moral proposition we should dislike it. No studies had ever tested this hypothesis until the authors’ experiments revealed the relation⁵⁵.

The experiments contrasted moral propositions, such as:

- (35) Torturing children should be a capital crime.
Prison should re-educate criminals.

with factual propositions, such as:

L.J. Skitka, D.C. Wisneski, *Moral Conviction and Emotion*, in «Emotion Review», 3, 2011 pp. 328-330; T. Wheatley, J. Haidt, *Hypnotic Disgust Makes Moral Judgments More Severe*, in «Psychological Science», 16, 2005, 780-784; L. Zhang, M. Kong, Z. Li, *Emotion Regulation Difficulties and Moral Judgment in Different Domains: The Mediation of Emotional Valence and Arousal*, in «Personality and Individual Differences», 109, 2017, pp. 56-60.

⁵¹ D. Westen, P.S. Blagov, K. Harenski, C. Kiltz, S. Hamann, *Neural Bases of Motivated Reasoning: an fMRI Study of Emotional Constraints on Partisan Political Judgment in the 2004 U.S. Presidential Election*, in «Journal of Cognitive Neuroscience», 18, 11, 2006, pp. 1947-1958; C. Tavis, E. Aronson, *Mistakes Were Made (But Not by Me): Why We Justify Foolish Beliefs, Bad Decisions, and Hurtful Acts*, Orlando (FL), Harcourt, 2007.

⁵² F.P. Ramsey, *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, London, Humanities Press, 190, pp. 145-163 (originally published in 1929).

⁵³ B. de Finetti, *La prevision: Ses lois logiques, ses sources subjectives*, in «Annales de l’Institut H. Poincaré», 7, 1937, pp. 1-68 (translated into English in H.E. Kyburg Jr, H.E. Smokler, eds., *Studies in Subjective Probability*, New York, Wiley, 1964).

⁵⁴ S. Khemlani, M. Lotstein, P.N. Johnson-Laird, *Naive Probability: Model-Based estimates of unique events*, in «Cognitive Science», 39, 2003, pp. 1216-1258.

⁵⁵ M. Bucciarelli, P.N. Johnson-Laird, *Emotions and Beliefs About Morality Can Change One Other*, 2019, under submission.

- (36) Governments who invest in culture are wise.
Religions are the most common cause of wars.

The participants had two tasks. They had to rate their degree of belief in each proposition on a five point scale from *impossible* to *certain*, and they had to rate their emotion about each proposition on a five point scale from *I loathe it* to *I love it*. In one experiment, two different groups of participants carried out the two tasks; in other experiments, the participants carried out both tasks but in counter-balanced orders. The results were always the same: there was no reliable correlation between emotion and belief for factual assertions, but a robust correlation between them for moral assertions.

The most striking experiment was one that contrasted moral and factual assertions that differed in only one word. For example, one of the moral assertions was:

- (37) Immigrants who do not find a job ought to go back to their country.

Its factual counterpart was:

- (38) Immigrants who do not find a job tend to go back to their country.

Each participant saw only one version of 30 such pairs, but 15 moral and 15 factual assertions. The experiment counterbalanced the order in which the participants rated their emotions and their strengths of belief. Figure 1 shows the mean ratings of belief and emotion for the 30 moral assertions. The strong correlation between the two is obvious: if you know the participants' strength of belief in a proposition, you can predict their emotional reaction to it with reasonable accuracy, and vice versa. Kendall's tau is a measure of the correlation between two rank orders, and it ranges from -1 (a perfect negative correlation) through 0 (no correlation whatsoever) to +1 (a perfect positive correlation). Its value for the data in Figure 1 was 0.82, a highly robust positive correlation. Figure 2 shows the mean ratings of belief and emotion for the 30 factual assertions. As Figure 2 makes clear, there was no reliable correlation ($\tau = .23$) between the two ratings for the factual assertions.

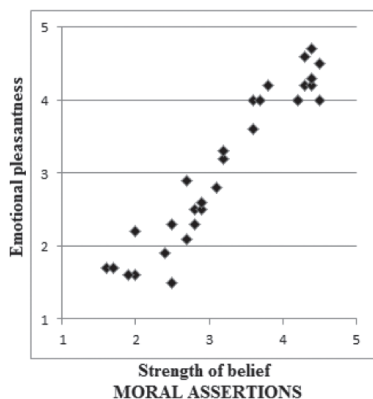


Figure 1. The participants' mean ratings of their emotional reactions (from loathing to loving) to 30 moral assertions and their strength of belief (from impossibility to certainty) in them.

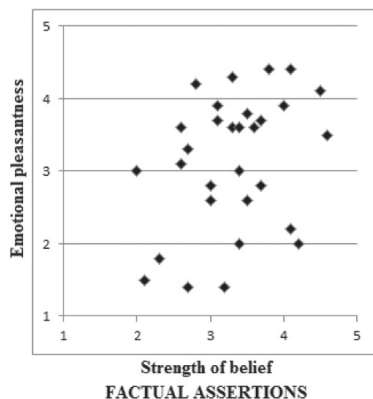


Figure 2. The participants' mean ratings of their emotional reactions (from loathing to loving) to 30 factual assertions and their strength of belief (from impossibility to certainty) in them.

Unlike factual beliefs, deontic propositions are not open to empirical test – they concern things unseen. Another real difference exists between moral and factual assertions. Both of them can elicit strong positive and strong negative emotions, and both of them can elicit strong beliefs and strong disbeliefs. Where they differ is that these two factors – emotion and belief – run in parallel only for moral propositions. The correlation tells us nothing about the causal relations between emotions and beliefs. The model theory's principle of independent systems allows that emotions can affect degrees of belief, and that degrees of belief can influence emotions. An ex-

periment corroborated the first of these causal relations. It changed the participants' emotions about moral propositions: they had to recall pleasant or unpleasant episodes in their lives pertinent to moral propositions. A parallel change occurred in the strength of their beliefs in these propositions: pleasant memories increased beliefs, and unpleasant memories decreased them⁵⁶. The changes were small, but ran in parallel together in a reliable way. This result also corroborates the socio-intuitionist theory⁵⁷, but it is inexplicable for a moral grammar⁵⁸. Likewise, another experiment corroborated the second of the model theory's causal claims. It changed the participants' degrees of belief in moral propositions: they had to create their own reasons for believing or disbelieving the moral propositions. A parallel change occurred in the strength of the emotions they elicited in the participants; reasons to believe increased positive emotions, and reasons to disbelieve increased negative emotions⁵⁹. This result corroborates the moral grammar theories but it is inexplicable for the socio-intuitionist theory. Over all, the results of the two experiments are contrary to both moral grammar and socio-intuitionist theories, but they are consistent with the assumption that emotions and reasoning depend on parallel systems that interact one another. An unanswered question, as yet, is whether deontic propositions that are not about morality can also elicit correlations between beliefs and emotions.

4. *Conclusions*

Our reasoning about deontic topics relies on mental models of what is possible in a context of relevant principles. The decisive evidence for this hypothesis is our susceptibility to illusory inferences about what is permissible and what is impermissible. Alternative theories offer no explanation for these illusions. Contrary to Hume and Haidt, we do reason to make moral judgments. Our inferences may be rapid intuitions, but they can also be slow deliberations as we try to resolve a dilemma. The occurrence of both sorts of rea-

⁵⁶ M. Bucciarelli, P.N. Johnson-Laird, *Emotions and Beliefs*, cit.

⁵⁷ R.A. Shweder, J. Haidt, *The Cultural Psychology of the Emotions: Ancient and New*, in M. Lewis, J. Haviland, eds., *Handbook of Emotions*, New York, NY, Guilford, 2000, pp. 397-414; J.Q. Wilson, *The Moral Sense*, New York (NY), Free Press, 1993.

⁵⁸ D.M. Hauser, *Moral Mind*, cit.; D.M. Hauser, *The Liver and the Moral Organ*, in «Social Cognitive and Affective Neuroscience», 1, 2006, pp. 214-220; J. Mikhail, *Elements of Moral Cognition*, Cambridge (MA), Harvard University Press, 2011.

⁵⁹ M. Bucciarelli, P.N. Johnson-Laird, *Emotions and Beliefs*, cit.

soning hardly squares with the idea that moral judgments derive from a grammar. The innate system of basic emotions, which humans inherited from social mammals, relies on primitive cognitions far too crude to determine that a person who causes a morally bad outcome is more blameworthy than a person who only enables it. The required inference is so subtle that some legal scholars have not recognized that the distinction is grounded in meanings. Yet, a striking relation exists between moral propositions and emotions – a relation that moral grammars also cannot explain. We love those moral propositions that we believe, and hate those that we disbelieve. No such relation exists for factual propositions. As the ancient Greek philosopher Epicurus remarked: pleasure is the measure of what is good. His claim is now backed by evidence.