Review

Cell PRESS

The acquisition of Boolean concepts

Geoffrey P. Goodwin[1] and Philip N. Johnson-Laird[2,3]

[1] Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Department of Psychology, Princeton University, Princeton, NJ 08544, USA
[3] Department of Psychology, New York University, New York, NY 10003, USA

Boolean relations, such as *and*, *or*, and *not*, are a fundamental way to create new concepts out of old. Classic psychological studies showed that such concepts differed in how difficult they were to learn, but did not explain the source of these differences. Recent theories have reinvigorated the field with explanations ranging from the complexity of minimal descriptions of a concept to the relative invariance of its different instances. We review these theories and argue that the simplest explanation – the number of mental models required to represent a concept – provides a powerful account. However, no existing theory explains the process in full, such as how individuals spontaneously describe concepts.

Introduction

Many everyday concepts combine existing elements using Boolean relations: a 'cousin' is a child of an uncle *or* aunt, and 'beer' is an alcoholic beverage usually made from malted cereal grain *and* flavored with hops, *and* brewed by slow fermentation. Laws and technical concepts also depend on Boolean relations: in baseball, a 'ball' is a pitch that does *not* enter the strike zone *and* is *not* struck at by the batter. Likewise, most of the diagnostic criteria in the *Diagnostic and Statistic Manual of Mental Disorders* (DSM-IV-TR) [1] rely on Boolean relations. What are Boolean relations, however? The nineteenth century English logician George Boole used them in his eponymous algebra to form new sets (or, as he wrote, 'classes') out of old, and they consist of relations corresponding to *and* (intersection), *or* (union), *not* (complement), and those further relations that can be defined in terms of them [2]. They have been applied to many domains, from the design of electronic circuits to the axiomatization of the probability calculus, which Boole anticipated in also interpreting his formulas as numerical probabilities.

A longstanding goal for cognitive psychology is to understand how people acquire, represent, and use concepts. The pioneering studies examined Boolean concepts, that is, those defined solely in terms of Boolean relations of such properties as color, size, and shape. They showed that concepts depending on *and* are easier to learn than those depending on *or* [3–5]. Neisser and Weene [6] proposed that the difficulty of acquiring a Boolean concept depends on the length of its 'minimal descriptions', that is, the shortest Boolean descriptions of all and only its instances. This account, however, did not explain a puzzling finding – the robust trend in difficulty observed by Shepard, Hovland, and Jenkins [7] over six different Boolean concepts (Box 1). We refer to this phenomenon as the 'Shepard trend'.

Research on Boolean concepts waned after the mid-1960s, leaving the Shepard trend unexplained, as psychologists turned to prototypes [8–12], exemplars of concepts, namely, their instances [13,14], and the role of theoretical knowledge in their acquisition [15,16]. Yet, Boolean relations play a role even in prototypes. A prototypical dog is one that has all the attributes in a set of defaults: it has four legs, *and* a tail, *and* it barks, etc. Recently, psychologists have renewed their study of Boolean concepts and have developed better accounts of their acquisition. The goal of unifying these accounts with a broader understanding of cognition has pulled theorists in different directions – towards algorithmic explanations of attention and working memory, and towards explanations of reasoning.

Recent accounts of the Shepard trend

Early attempts to explain the trend [6,17–20] (but cf. [21]) were not as successful as recent theories. One such recent theory is ALCOVE, a three-layer feed-forward connectionist network [22], which was developed from earlier exemplar theories [13,14]. It represents individual exemplars as points in a multi-dimensional space. Another such theory is SUSTAIN, a connectionist network that represents concepts as clusters akin to prototypes in a multi-dimensional space [23]. The number of clusters predicts the difficulty of acquiring concepts, including those in the Shepard trend. However, these connectionist networks are not superior to alternative theories based on very different mechanisms. For example, RULEX represents Boolean concepts as combinations of rules and their exceptions [24]. It requires less memory than ALCOVE and it predicts the difficulty of learning concepts from the number of variables and the number of exceptions that it needs to store. Latterly, theorists have built on exemplar theories to develop accounts based on rules in first-order logic that are acquired according to Bayesian probabilities [25]. All these theories predict the Shepard trend. It is a useful initial benchmark, but it fails to discriminate among rival accounts.

The field needed a larger body of robust results, and Feldman [26] provided them for the acquisition of a battery of 76 Boolean concepts, which were instantiated as, 'amoebas' defined by binary properties, such as shape, size, and

Corresponding authors: Goodwin, G.P. (ggoodwin@psych.upenn.edu); Johnson-Laird, P.N. (phil@princeton.edu).

1364-6613/$ – see front matter
 http://dx.doi.org/10.1016/j.tics.2013.01.007

Box 1. The Shepard *et al.* [7] task and its principal results

Shepard, Hovland, and Jenkins [7] investigated the acquisition of six different sorts of concept based on three binary variables. Each concept had four instances and four non-instances in the eight possibilities. Figure I presents the six concepts: each dimension represents a binary variable and the instances of the six concepts are shown as black blobs (see also Table 1). In one study, the instances were geometrical shapes; in another study, they were everyday objects. In some conditions, the variables concerned a single entity, such as its shape, size, and color. In other conditions, they concerned separate entities. In the first study, the instances and non-instances of each concept were presented sequentially and participants at first guessed whether or not a particular entity was an instance and gradually acquired the concept. In the two subsequent studies, the instances and non-instances were presented simultaneously and identified as such at the beginning of the task. The participants had to remember the concept and to formulate a rule for it. The experiments also varied the nature of the responses: the stimuli were variously classified as in 'A' and 'B' groups, 'plus' and 'minus' groups, '1' and '2' groups, and so on. The dependent variables included the number of errors during learning, latencies, and, where relevant, the accuracy of the participants' descriptions.

Despite all of the differences in method, the results were highly consistent, and showed the following partial trend from easiest to hardest: I < II < III, IV, V < VI. This trend in difficulty has been replicated many times (e.g., [55–57]).

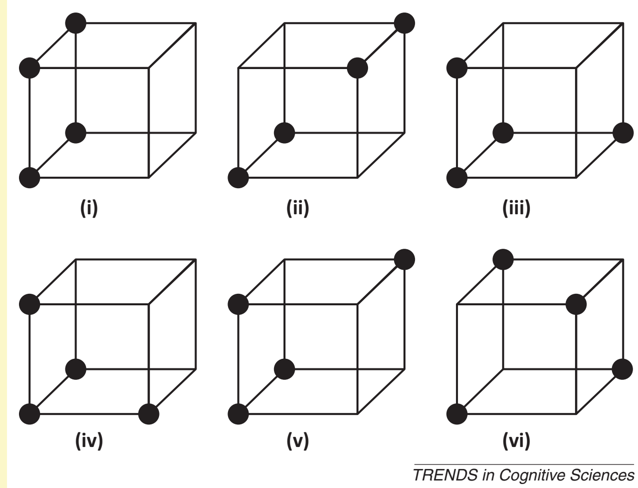

Figure I. Examples of the six sorts of concept used by Shepard *et al.* [7]. The black blobs represent the instances of the concepts.

shading. His data did not replicate the precise Shepard trend, but they provide a powerful challenge to theories. We refer to them as the 'Feldman dataset'. His research [26,27] and other studies [28–32] revived interest in concepts.

Minimal complexity

There are infinitely many descriptions of the same Boolean concept, but most contain redundant clauses. For instance, the following two descriptions describe the same concept:

(a and b) or (a and not b) or (not a and b)

and:

(a or b)

The first is in 'disjunctive normal form' and the second is a minimal description, that is, no shorter description of the concept is possible. Feldman proposed that the length of minimal descriptions should predict the difficulty of learning concepts [26]. However, because *a and b* is easier than *a or b*, he made a further 'parity' assumption: concepts with fewer instances than non-instances should be easier to learn than those with fewer non-instance than instances (cf. [6,7]). The resulting theory appeared to account for the Shepard trend and for just over 50% of the variance in the Feldman dataset. These results seemed to be a major advance; but there were problems. The descriptions in [26], as several authors pointed out [33–36], were not truly minimal, whereas the correct minimal descriptions make much less successful predictions. Likewise, no good reason exists for restricting minimal descriptions to *not*, *and*, and *or*. Why not allow *or else* (exclusive disjunctions)? But, if they are allowed, the predictions are much less successful. The process for constructing minimal descriptions is too complex to be psychologically plausible, no good evidence existed to show that participants construct them, and, as we report below, they seldom do construct them. Feldman's theory explained an impressive amount of variance, but it left room for improvement, in part because 'parity' explained almost 20% of the variance. Hence, there has been a recent flurry of theorizing, to which we now turn.

Current theories

Feldman devised a new theory of algebraic complexity, in which Boolean concepts are decomposed into a set of simpler underlying formulas [37]. Consider, for example, the concept with the following instances, where '¬' denotes the absence of a property:

¬ a	¬ b	¬ c
¬ a	¬ b	c
a	¬ b	¬ c
a	¬ b	c

These instances can be decomposed into the conjunction (the Cartesian product) of the alternatives: ¬ a ¬ b, a ¬ b with the alternatives ¬ c, c. A weighted average of the complexity of these underlying components, which depends on the number of variables in them (akin to their minimal complexity), yields the concept's overall algebraic complexity. There are advantages in decomposition [38] and in the general goal of simplicity [39–41]. Feldman's theory copes not just with binary variables, but also with variables that have multiple values. But, its computations are complex, and its accounts for what is computed rather than how it is computed. Nevertheless, its single parameter accounts for approximately 50% of the variance in the Feldman dataset.

Vigo [42] developed a theory of Boolean concepts in terms of the degree to which their instances are similar to one another, that is, their structural invariance. The computation also takes into account the number of instances in a concept, because the fewer they are, the easier the concept should be to acquire. The structural complexity of a concept is therefore inversely proportional to its degree of invariance and directly proportional to its cardinality. The computations underlying this theory are also complex and it too gives no account of mental processes. Yet, once again, the fit of this theory is impressive, accounting for approximately 42% of the variance in the Feldman dataset.

Both algebraic complexity and structural invariance account for the Feldman data set better than either ALCOVE [22] or SUSTAIN [23] does. And no simple way exists to extend RULEX [24] to predict the Feldman dataset. Yet, neither algebraic complexity nor structural invariance accounts for the mental representation of concepts or for the processes underlying their acquisition – the computations of complexity and invariance are likely to be beyond most individuals. Hence, we turn to a theory that aims to explain mental representations and processes.

The model theory of concepts

The theory of mental models presupposes that the mind is neither a logical nor a probabilistic device. It makes simulations. The theory applies to reasoning in general and it postulates that reasoners try to envisage the possibilities to which premises refer and draw conclusions that hold in them [43,44]. As the theory predicts, the more models of possibilities that individuals have to represent, the harder an inference is (e.g., [45]). The instances of a Boolean concept can each be represented in a mental model. However, as individuals acquire the concept, they can reduce the number of these models, by eliminating those variables that are irrelevant given the values of other variables [33]. Box 2 describes this idea in detail.

The number of models of a concept that results from the elimination of irrelevant variables predicts the difficulty of acquiring the concept. Current evidence gives the theory a slight edge over rival accounts. Table 1 summarizes its predictions and those of minimal descriptions, algebraic complexity, and structural complexity for the Shepard trend. The model theory predicts these results approximately as well as do rival theories. However, for the Feldman dataset, the number of mental models accounts for 57% of the variance in the difficulty of acquiring them, which is an improvement of approximately 8% more than algebraic complexity [37] and of 15% more than structural invariance [42]. Nevertheless, when models and algebraic complexity were entered into a regression as simultaneous predictors of accuracy, both accounted for unique variance, revealing that algebraic complexity may be capturing something that models do not, and vice versa [33].

Box 2. The simplification of models of concepts

The process of simplifying mental models of the instances of concepts is implemented in a computer program written in LISP. It takes descriptions of Boolean concepts as input, and outputs a set of simplified models that represent the concept [33]. It works by eliminating irrelevant variables. Consider the following instances of the symptoms of a disease, where '¬' denotes the absence of a symptom:

Fever	Headache	Rash
Fever	¬ Headache	Rash

Here, 'fever' and 'rash' occur with both with and without 'headache', and so this variable is irrelevant to the concept. Any case in which both fever and rash occur is an instance of the disease whether or not the patient has a headache. The models are accordingly simplified to:

Fever	Rash

The burden on working memory is alleviated, because the concept is transformed from two models to one.

In other cases, more radical simplifications occur. Consider concept I from Shepard's six concepts [8]:

Fever	Headache	Rash
Fever	Headache	¬ Rash
Fever	¬ Headache	Rash
Fever	¬ Headache	¬ Rash

The variables of 'headache' and 'rash' can be eliminated, because all four pairs of their possible joint values occur together with 'fever', and so the result is:

Fever

For some concepts, successive simplifications can be made – for example, concept III from Shepard *et al.* [7] is:

Fever	¬ Headache	Rash
Fever	¬ Headache	¬ Rash
¬ Fever	Headache	Rash
¬ Fever	Headache	¬ Rash

When 'fever' and '¬ headache' co-occur, the variable concerning 'rash' is irrelevant, and can be eliminated to yield:

Fever	¬ Headache	
¬ Fever	Headache	Rash
¬ Fever	Headache	¬ Rash

When '¬ fever' and 'headache' co-occur, the variable concerning 'rash' is again irrelevant, and can be eliminated to yield:

Fever	¬ Headache
¬ Fever	Headache

When alternative simplifications of the same concept are possible, the computer program searches for and returns a simplification with the fewest number of models.

Descriptions of concepts

Studies of concept acquisition vary in methodological details, but a common thread is that the participants' task is to identify the instances and non-instances of concepts. Performance is objective and easy to quantify, but it can be less revealing than participants' descriptions of concepts. With few exceptions [7], this task has not been used for Boolean concepts. However, when participants had to acquire concepts defined in terms of the positions of three switches that controlled a light, their descriptions of the concepts revealed several striking phenomena [33]:

- Only 2% of descriptions were minimal Boolean descriptions and the correct minimal descriptions failed to predict the difficulty of the task.
- The majority of descriptions were in the form of disjunctions, which is in accord with representations in the form of alternative mental models.
- Many descriptions used language outside the scope of Boolean relations, as illustrated in these uses of other relations, quantifiers, and numbers:

> The light will go on when switches one and two are in the same position.
>
> If all of the switches are turned to the right, the light will be off, otherwise it is on.
>
> If the sum of the switches turned to the right is more than two, the light will be on.

A second experiment replicated these phenomena, and showed that language outside Boolean relations can make an otherwise difficult problem easy. Every theory predicts that the following concept should be difficult, but this description shows why it was easy:

Table 1. The instances of the six Shepard *et al.* [7] concepts in Figure 1, their simplified mental models [33], correct minimal descriptions revised from [26], algebraic complexity [37], and structural complexity [42]. ('¬' denotes the absence of a property.)

Concept number	Instances of the concept	Simplified mental models	Minimal descriptions	Algebraic complexity	Structural complexity
I	¬ a b c ¬ a b ¬ c ¬ a ¬ b c ¬ a ¬ b ¬ c	¬ a	*not* a (1)	-.50	1.66
II	a b c a b ¬ c ¬ a ¬ b c ¬ a ¬ b ¬ c	a b ¬ a ¬ b	(a *and* b) *or* (*not* a *and not* b) (4)	0.00	2.00
III	a ¬ b c ¬ a b ¬ c ¬ a ¬ b c ¬ a ¬ b ¬ c	¬ a ¬ c ¬ b c	(*not* a *and not* c) *or* (*not* b *and* c) (4)	0.00	2.14
IV	a ¬ b ¬ c ¬ a b ¬ c ¬ a ¬ b c ¬ a ¬ b ¬ c	¬ a b ¬ c ¬ a ¬ b c ¬ b ¬ c	(*not* c *or* (*not* a *and not* b)) *and* (*not* a *or not* b) (5)	0.00	2.14
V	a b c ¬ a b ¬ c ¬ a ¬ b c ¬ a ¬ b ¬ c	a b c ¬ a b ¬ c ¬ a ¬ b	(*not* a *and not* (b *and* c)) *or* (a *and* (b *and* c)) (6)	.50	2.34
VI	a b ¬ c a ¬ b c ¬ a b c ¬ a ¬ b ¬ c	a b ¬ c a ¬ b c ¬ a b c ¬ a ¬ b ¬ c	(a *and* ((*not* b *and* c) *or* (b *and not* c))) *or* (*not* a *and* ((*not* b *and not* c) *or* (b *and* c))) (10)	2.00	4.00

The light comes on in all cases when the switches are all on or all off.

No existing theory provides an adequate account of how individuals spontaneously describe concepts. Kemp [46] attempts to characterize the complete space of possible sorts of concepts, and he takes objects, properties, and relations as primitives, and first-order logic as the language to represent concepts. (First-order logic allows quantified variables to range only over entities, as opposed to their properties.) The primitives combined in different ways yield different sorts of concept. However, like earlier theories [26,40], his account postulates that the difficulty of acquiring a concept depends on the length of its minimal description in the representation language. Logic provides an account of the mapping of models into descriptions using quantifiers and relations, but so too does set theory. And set theory, which underlies Boole's algebra, has the advantage that it captures all quantifiers in a unified way, as Montague showed [47], including such quantifiers as 'more than half the electorate', which cannot be defined with the quantifiers of first-order logic [48]. However, as Boole realized, a quantified assertion states a relation between sets [2]. This treatment accommodates all quantifiers and it is how quantifiers are represented according to the model theory [49].

Models are similar to logical descriptions of concepts in disjunctive normal form (see above). So, is there a crucial phenomenon that corroborates the model theory? Indeed, there is such a phenomenon, and it occurs when individuals interpret descriptions of concepts. The model theory postulates a principle of truth: mental models represent what is true, not what is false. The principle reduces the load on working memory, but it has an unexpected consequence discovered only from the theory's computer implementation [50]. It predicts the occurrence of systematic

Box 3. Conceptual illusions

A corollary of the principle of truth for concepts is that mental models represent only the instances of a concept and that for each instance they represent only those properties, or their absence, that the description ascribes to the instance [54]. As an example, consider this description of a concept:

It's red and it's square, or else it's red

The mental models of the concept are as follows:

Red Square (the first clause of the disjunction is true)
Red (the second clause of the disjunction is true)

Given the description of the concept, most participants listed these two instances [54]. However, they were wrong, and wrong about other problems, because mental models represent only what is true. Fully explicit models also represent what is false. When the first clause of the disjunction is true, there is one possible instance:

Red Square

but the force of *or else* is that in this case the second clause is false – it isn't red – and so the first clause must be false (and the second clause true):

Red ¬ Square

A crucial feature of such concepts is that the mental models contain a model of an 'illusory' instance, one that does not correspond to a real instance. Individuals err on such concepts, even those that do not depend on *or else*, but they perform accurately on control problems with similar descriptions, similar numbers of models, but no models of illusory instances [54,58].

The model theory postulates that the interpretation of relations can be modulated by semantics and knowledge. Studies of deductive reasoning have corroborated such effects in several domains [59–61]. One effect is to block the construction of a model, and so an experiment on concepts [54] used contents that should have this effect on illusory models, such as:

It's red and it's green, or else it's red.

Individuals know that the first clause cannot be true (assuming uniform colors), and so the second clause is true. There is accordingly only one possibility:

Red ¬ Green

An experiment showed that the original materials elicited illusory instances, whereas modulated materials of this sort elicited reliably more correct instances [54].

Box 4. Current problems and questions for future research

- Sets of a fixed number of mental models are not necessarily equally easy to acquire. The ease of decomposing models into conjunctions of their components probably correlates with algebraic complexity [37]. The distance apart of models in conceptual space – perhaps using their Levenshtein distances [62] (but cf. [63,64]) – probably correlates inversely with invariance [42]. So, could a rapprochement among current theories explain the unaccounted variance (approximately 50%) in Boolean concept acquisition?
- Boolean relations do not exhaust the machinery for building new concepts out of old. How is their acquisition integrated with learning other sorts of relation that also underlie many concepts [65–70], and learning those concepts, such as 'elevator' and 'steam engine', that seem likely to depend on mental simulations?
- The relative difficulty of acquiring concepts appears to change from those based on the presence or absence of visual properties [7,26] to those based on switch positions [33]. What are the critical variables underlying these differences? Do they explain the discrepancies from the Shepard trend?
- Many concepts, perhaps most, are acquired from descriptions and so the interpretation and formulation of such descriptions are a major part of any theory of concepts. How can individuals' spontaneous use of quantifiers and other non-Boolean expressions best be accounted for? And under what conditions are such expressions most likely to be generated? How much do they depend on similarities in the way different variables that define a concept are instantiated?
- The revival in the study of Boolean concepts is due to a theory about their simplest possible descriptions [26]. The model theory suggests instead that the crucial variable is the simplest possible set of models of a Boolean concept. Does the simplest description have a role to play in an expanded linguistic account that includes quantifiers and numbers?
- Classic studies [3] investigated the role of strategies in concept acquisition, and they are important in reasoning (e.g., [71]). What role do they and order of presentation (e.g., [72]) play in Boolean concept acquisition?
- A computer system that acquires novel concepts calls for explicit theories of many processes, including attention, memory, and the sort of machinery in the model theory. How can such a theory be best integrated within accounts of working memory (cf. [73–75]), communicative efficiency [76], and a general mental architecture, such as Anderson's ACT-R [77,78]?

fallacies, which are compelling enough to be cognitive illusions. They have been corroborated in deductions [51], evaluations of the consistency of sets of assertions [52], and reasoning of many other sorts [53]. Illusory concepts also occur [54], and they are described in Box 3. No other existing theory of Boolean concepts predicts their occurrence.

Concluding remarks

After a hiatus, and thanks to the work of Feldman [26] and others, the study of Boolean concept acquisition is flourishing. Several recent theories have predicted with some success the difficulty of learning concepts. They include accounts based on the minimal descriptions of concepts [26,46], the decomposition of their formulas into conjunctive components [37], the assessment of the degree to which the instances of a concept are invariant [42], and the simplification of their exemplars into more parsimonious mental models [33,54]. As authors of the model theory, we believe that it has four main advantages over alternative accounts:

- It is simple enough to give a plausible account both of mental representation of concepts and of the mental processes underlying their acquisition.
- It provides a better account than other theories of the acquisition of the 76 concepts in Feldman's dataset [26].
- It predicts two novel phenomena (Box 3).
- Its basis in simulation and set theory allows it to generalize to concepts based on relations, including temporal and causal ones, and to all sorts of quantifiers.

Nonetheless, we concede that it leaves much unaccounted for, and Box 4 summarizes the main lacunae and goals for future research.

Acknowledgements

The research was supported in part by Grant SES 0844851 to P.N.J-L. from the National Science Foundation to study deductive and probabilistic reasoning.

References

1 American Psychiatric Association (2000) *Diagnostic and Statistical Manual of Mental Disorders* (4th edn., text revision) – DSM-IV-TR, American Psychiatric Association.
2 Boole, G. (1854) *An Investigation of the Laws of Thought*, Macmillan
3 Bruner, J.S. *et al.* (1956) *A Study of Thinking*, Wiley
4 Conant, M.B. and Trabasso, T. (1964) Conjunctive and disjunctive concept formation under equal-information conditions. *J. Exp. Psychol.* 67, 250–255
5 Haygood, R.C. and Bourne, L.E., Jr (1965) Attribute and rule learning aspects of conceptual behavior. *Psychol. Rev.* 72, 175–195
6 Neisser, U. and Weene, P. (1962) Hierarchies in concept attainment. *J. Exp. Psychol.* 64, 640–645
7 Shepard, R.N. *et al.* (1961) Learning and memorization of classifications. *Psychol. Monogr. Gen. Appl.* 75, 1–42
8 Hampton, J.A. (1979) Polymorphous concepts in semantic memory. *J. Verb. Learn. Verb. Behav.* 18, 441–461
9 Posner, M.I. and Keele, S.W. (1968) On the genesis of abstract ideas. *J. Exp. Psychol.* 77, 353–363
10 Posner, M.I. and Keele, S.W. (1970) Retention of abstract ideas. *J. Exp. Psychol.* 83, 304–308
11 Rosch, E. and Mervis, C.B. (1975) Family resemblance: studies in the internal structure of categories. *Cogn. Psychol.* 7, 573–605
12 Smith, E.E. and Medin, D.L. (1981) *Categories and Concepts*, Harvard University Press
13 Medin, D.L. and Schaffer, M.M. (1978) Context theory of classification learning. *Psychol. Rev.* 85, 207–238
14 Nosofsky, R.M. (1986) Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115, 39–57
15 Keil, F.C. (1989) *Concepts, Kinds, and Cognitive Development*, MIT Press
16 Murphy, G. and Medin, D.L. (1985) The role of theories in conceptual coherence. *Psychol. Rev.* 92, 289–316
17 Garner, W.R. (1962) *Uncertainty and Structure as Psychological Concepts*, Wiley
18 Hovland, C.I. and Weiss, W. (1953) Transmission of information concerning concepts through positive and negative instances. *J. Exp. Psychol.* 45, 175–182
19 Hunt, E.B. (1962) *Concept Learning: An Information Processing Problem*, Wiley
20 Hunt, E.B. *et al.* (1966) *Experiments in Induction*, Academic Press
21 Lafond, D. *et al.* (2009) Decision-tree models of categorization response times, choice proportions, and typicality judgments. *Psychol. Rev.* 116, 833–855
22 Kruschke, J.K. (1992) ALCOVE: an exemplar-based connectionist model of category learning. *Psychol. Rev.* 99, 22–44
23 Love, B.C. *et al.* (2004) SUSTAIN: A network model of category learning. *Psychol. Rev.* 111, 309–332
24 Nosofsky, R.M. *et al.* (1994) Rule-plus-exception model of classification learning. *Psychol. Rev.* 101, 53–79
25 Goodman, N.D. *et al.* (2008) A rational analysis of rule-based concept learning. *Cogn. Sci.* 32, 108–154

26 Feldman, J. (2000) Minimization of Boolean complexity in human concept learning. *Nature* 407, 630–633
27 Feldman, J. (2003) A catalog of Boolean concepts. *J. Math. Psychol.* 47, 75–89
28 Minda, J.P. and Smith, J.D. (2002) Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 275–292
29 Nosofsky, R.M. and Palmeri, T.J. (1997) An exemplar-based random walk model of speeded categorization. *Psychol. Rev.* 104, 266–300
30 Pothos, E.M. and Chater, N. (2002) A simplicity principle in unsupervised human categorization. *Cogn. Sci.* 26, 303–343
31 Smith, J.D. and Minda, J.P. (1998) Prototypes in the mist: the early epochs of category learning. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1411–1436
32 Smith, J.D. and Minda, J.P. (2000) Thirty categorization results in search of a model. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 3–27
33 Goodwin, G.P. and Johnson-Laird, P.N. (2011) Mental models of Boolean concepts. *Cogn. Psychol.* 63, 34–59
34 Lafond, D. *et al.* (2007) Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *J. Math. Psychol.* 51, 57–74
35 Mathy, F. and Bradmetz, J. (2004) A theory of the graceful complexification of concepts and their learnability. *Curr. Psychol. Cogn.* 22, 41–82
36 Vigo, R. (2006) A note on the complexity of Boolean concepts. *J. Math. Psychol.* 50, 501–510
37 Feldman, J. (2006) An algebra of human concept learning. *J. Math. Psychol.* 50, 339–368
38 Joshi, S. (2010) Concept learning by example decomposition. *J. Exp. Theor. Artif. Intell.* 2, 1–21
39 Chater, N. (1999) The search for simplicity: A fundamental cognitive principle? *Q. J. Exp. Psychol. Sec. A* 52, 273–302
40 Chater, N. and Vitányi, P. (2003) Simplicity: a unifying principle in cognitive science. *Trends Cogn. Sci.* 7, 19–22
41 Grünwald, P. (2007) *The Minimum Description Length Principle*, MIT Press
42 Vigo, R. (2009) Categorical invariance and structural complexity in human concept learning. *J. Math. Psychol.* 53, 203–221
43 Johnson-Laird, P.N. (2006) *How We Reason*, Oxford University Press
44 Johnson-Laird, P.N. (2010) Mental models and human reasoning. *Proc. Natl. Acad. Sci. U. S. A.* 107, 18243–18250
45 Byrne, R.M.J. and Johnson-Laird, P.N. (2009) 'If' and the problems of conditional reasoning. *Trends Cogn. Sci.* 13, 282–286
46 Kemp, C. (2012) Exploring the conceptual universe. *Psychol. Rev.* 119, 685–722
47 Montague, R. (1974) *Formal Philosophy: Selected Papers*, Yale University Press
48 Barwise, J. and Cooper, R. (1981) Generalized quantifiers and natural language. *Linguist. Philos.* 4, 159–219
49 Johnson-Laird, P.N. (1983) *Mental Models*, Cambridge University Press
50 Johnson-Laird, P.N. and Savary, F. (1999) Illusory inferences: a novel class of erroneous deductions. *Cognition* 71, 191–229
51 Khemlani, S. and Johnson-Laird, P.N. (2009) Disjunctive illusory inferences and how to eliminate them. *Mem. Cogn.* 37, 615–623
52 Johnson-Laird, P.N. *et al.* (2012) The consistency of disjunctive assertions. *Mem. Cogn.* 40, 769–778
53 Johnson-Laird, P.N. (2012) Inference in mental models. In *The Oxford Handbook of Thinking and Reasoning* (Holyoak, K. and Morrison, R., eds), pp. 134–154, Oxford University Press
54 Goodwin, G.P. and Johnson-Laird, P.N. (2010) Conceptual illusions. *Cognition* 114, 253–265
55 Nosofsky, R.M. *et al.* (1994) Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem. Cogn.* 22, 352–369
56 Love, B.C. (2002) Comparing supervised and unsupervised category learning. *Psychol. Bull. Rev.* 9, 829–835
57 Smith, J.D. *et al.* (2004) Category learning in rhesus monkeys: A study of the Shepard, Hovland, and Jenkins (1961) tasks. *J. Exp. Psychol. Gen.* 133, 398–414
58 Legrenzi, P. *et al.* (2003) Models of consistency. *Psychol. Sci.* 14, 131–137
59 Juhos, C. *et al.* (2012) Temporal and spatial relations in sentential reasoning. *Cogntion* 122, 393–404
60 Orenes, I. and Johnson-Laird, P.N. (2012) Logic, models, and paradoxical inferences. *Mind Lang.* 27, 357–377
61 Quelhas, A.C. *et al.* (2010) The modulation of conditional assertions and its effects on reasoning. *Q. J. Exp. Psychol.* 63, 1716–1739
62 Levenshtein, V. (1966) Binary codes capable of correcting deletions, insertions, and reversals. *Sov. Phys. Dokl.* 10, 707–710
63 Kornell, N. and Bjork, R.A. (2008) Learning concepts and categories: is spacing the 'enemy of induction'? *Psychol. Sci.* 19, 585–592
64 Wahlheim, C.N. *et al.* (2011) Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Mem. Cogn.* 39, 750–763
65 Doumas, L.A.A. *et al.* (2008) A theory of the discovery and predication of relational concepts. *Psychol. Rev.* 115, 1–43
66 Gentner, D. *et al.* (2011) Structure mapping and relational language support children's learning of relational categories. *Child Dev.* 82, 1173–1188
67 Lu, H. *et al.* (2012) Bayesian analogy with relational transformations. *Psychol. Rev.* 119, 617–648
68 Mathy, F. (2010) The long-term effect of relational information in classification learning. *Eur. J. Cogn. Psychol.* 22, 360–390
69 Mathy, F. and Bradmetz, J. (2011) An extended study of the nonindependence of stimulus properties in human classification learning. *Q. J. Exp. Psychol.* 64, 41–64
70 Sloutsky, V.M. (2010) From perceptual categories to concepts: what develops? *Cogn. Sci.* 34, 1244–1286
71 Van der Henst, J-B. *et al.* (2002) Strategies in sentential reasoning. *Cogn. Sci.* 26, 425–468
72 Mathy, F. and Feldman, J. (2009) A rule-based presentation order facilitates category learning. *Psychon. Bull. Rev.* 16, 1050–1057
73 Lewandowsky, S. (2011) Working memory capacity and categorization: Individual differences and modeling. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 720–738
74 Newell, B.R. *et al.* (2010) The dimensionality of perceptual category learning: A state-trace analysis. *Mem. Cogn.* 38, 563–581
75 Nosofsky, R.M. *et al.* (2011) Short-term memory scanning viewed as exemplar-based categorization. *Psychol. Rev.* 118, 280–315
76 Kemp, C. and Regier, T. (2012) Kinship categories across languages reflect general communicative principles. *Science* 336, 1049–1054
77 Fific, M. *et al.* (2010) Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk and decision-bound approaches. *Psychol. Rev.* 117, 309–348
78 Khemlani, S. and Trafton, J.G. (2012) mReactr: A computational theory of deductive reasoning. In *Proc. 34th Annu. Conf. Cogn. Sci. Soc.* (Miyake, N. *et al.*, eds), pp. 581–586, Erlbaum