5.4 Defeasible Reasoning and Belief Revision in Psychology

Lupita Estefania Gazzo Castañeda and Markus Knauff

Summary

How do people change their mind in the light of new information? In defeasible reasoning, people withdraw previously drawn *inferences* (conclusions) in light of new evidence. In belief revision, people abandon prior *beliefs* (premises) to accommodate new information. Cognitive research aims at developing *descriptive* theories of these cognitive functions, without making strong *normative* claims. However, such normative claims are more or less implicit in many cognitive theories of human reasoning. This chapter summarizes the experimental state of the art in the field and then presents the most important psychological theories of defeasible reasoning and belief revision. We also discuss whether both cognitive processes are really so different from a cognitive (descriptive) point of view. Both cognitive processes seem to rely on the availability of disablers, the facts that prevent people from believing the conclusions or premises of a reasoning problem.

1. Defeasible Reasoning and Belief Revision in Human Reasoning

A fundamental prerequisite of rationality is that humans are willing to change their mind in light of new evidence. For instance, suppose that Jack has learned that if he goes to bed late, then he will be tired the next day. So, if he is going to bed late, he can conclude that he will be tired the next day. But if he finds out that the next day is a holiday, then it makes sense for Jack to change this conclusion. He can conclude that he will not be tired because he is able to sleep in. That is, new information can make people reject previous conclusions that would have been drawn otherwise. Sometimes, however, the information we get is more than an exception to a rule. New information can also conflict with our beliefs. Imagine, for instance, that Jack usually has problems falling asleep. So the psychotherapist tells him that if he meditates in the evening, he will fall asleep more easily. Following the advice, Jack meditates in the evening. But he is still not able to fall asleep. This is not what he expected. Therefore, Jack has to revise his prior beliefs and conclude either that the psychotherapist was wrong or that he meditated inappropriately.

These two cases of reasoning are called defeasible reasoning and belief revision, respectively. In *defeasible reasoning*, people withdraw previously drawn conclusions in light of new evidence. Jack withdraws the conclusion of being tired after going to bed late because he knows that he is able to sleep in. This does not mean that he no longer believes that going late to bed makes him tired. In fact, in defeasible reasoning, one still believes in the prior propositions (basic beliefs), but one knows that in light of specific circumstances, some conclusions (derived beliefs) are not suitable anymore. This is different in *belief revision*, in which new information is in conflict with one's prior beliefs. The fact that Jack is not able to fall asleep is inconsistent with what the psychotherapist told him. This inconsistency between new evidence and prior beliefs can only be solved by revising these prior beliefs. In other words, in belief revision, it is not the conclusion that is changed; rather, one has to change prior beliefs in order to reach consistency. In the psychology of reasoning, these prior beliefs are often given by a set of premises. The conclusion is usually the belief that follows from these premises.

From a formal point of view, defeasible reasoning and belief revision are *nonmonotonic*, that is, conclusions can be retracted based on further evidence. Psychologically, that means that a person abandons accepted consequences from previous beliefs (Elio & Pelletier, 1997). The investigation of nonmonotonicity has its origins in philosophy and artificial intelligence (AI), in which many *normative* theories have been developed to describe how intelligent systems can rationally operate with incomplete information in changing worlds (for an overview, see Antoniou, 1997). In psychology, however, the investigation of defeasible reasoning and belief revision is still young. For a long time, psychologists concentrated only on deduction, which is monotonic, that is, a conclusion has to follow necessarily from its premises and cannot be altered by additional information. However, as our initial examples show, it is also sensible to change conclusions in light of new evidence. Because of that, psychologists have also started to develop several *descriptive* theories of human nonmonotonic reasoning and the psychological processes behind defeasible reasoning and belief revision. The aim of this chapter is to give an overview of this psychological research.

2. The Empirical Investigation of Defeasible Reasoning

Defeasibility is a central property of human reasoning. In everyday life, people are often forced to reason with uncertain and incomplete information and thus must often change previously drawn conclusions in light of new evidence. Most of the psychological research on defeasible reasoning focused on *conditional inferences* and the circumstances under which the conclusions from such inferences are withdrawn. There are two main experimental paradigms: the overt and the covert paradigm (Politzer & Bonnefon, 2006).

2.1 Defeasibility in the Overt Paradigm

The overt paradigm was introduced by Rumain, Connell, and Braine (1983) but made popular by Byrne (1989). There are usually two experimental conditions. In the simple condition, participants have to solve inferences consisting of a conditional premise of the form "If *p*, then *q*," a categorical premise containing *p*, *q*, or their negations and a conclusion that has to be evaluated. An example:

If Jack goes to bed late, then Jack will be tired.

Jack goes to bed late.

Jack will be tired.

In this case, the conclusion is true because this is the valid modus ponens (MP) inference: if *p* (the antecedent) is true, then *q* (the consequent) is also true. Another valid inference is modus tollens (MT: if *p* then *q*; not-*q*; therefore, not-*p*). By contrast, the other two inferences, *affirmation of the consequent* (AC: if *p* then *q*; *q*; therefore, *p*) and *denial of the antecedent* (DA: if *p* then *q*; not-*p*; therefore, not-*q*), are fallacies in classical logic because *p* is only sufficient, but not necessary, for *q*. Participants typically have no problem in recognizing MP inferences as logically valid, and they do it also for MT inferences, although to a lesser extent. The results regarding AC and DA are mixed and highly dependent on the task demands (Evans, Newstead, & Byrne, 1993).

In the extended condition, though, an additional premise is added to the inference. This additional premise can contain either a disabler or an alternative (Cummins, 1995; Cummins, Lubart, Alksnis, & Rist, 1991). *Disablers* are additional preconditions that prevent *p* from leading to *q*. They are thus relevant for MP and MT and can be presented either as concrete facts (e.g., Jack is able to sleep in) or as conditionals. An example:

If Jack goes to bed late, then he will be tired.

If Jack is able to sleep in, then he will not be tired.

Jack goes to bed late.

Jack will be tired.

Alternatives, by contrast, are alternative reasons for *q* that are not *p*. They are relevant for AC and DA inferences because they make it clear that *p* is not necessary for *q*. Alternatives can be also phrased either as facts (e.g., Jack has a disrupted sleep) or as conditionals. For example:

If Jack goes to bed late, then he will be tired.

If Jack has a disrupted sleep, then he will be tired.

Jack does not go to bed late.

Jack will not be tired.

The defeasibility of human reasoning is shown by comparing the acceptance rates for conclusions in the simple and the extended condition. This comparison usually shows that disablers and alternatives affect inferences differently. When disablers are presented, participants accept fewer MP and MT conclusions than without disablers (e.g., Byrne, 1989; Byrne, Espino, & Santamaria, 1999; De Neys, Schaeken, & d'Ydewalle, 2003b; Stevenson & Over, 1995). When alternatives are presented, participants accept fewer AC and DA conclusions than without alternatives (e.g., Byrne, 1989; De Neys et al., 2003b; Rumain et al., 1983).

2.2 Defeasibility in the Covert Paradigm

The covert paradigm was used for the first time by Cummins (Cummins et al., 1991). In the covert paradigm, alternatives and disablers are not presented explicitly as additional premises but are just implicitly present in the premises. For this, a first group of participants is confronted with a set of conditionals and asked to generate as many disablers and alternatives as they can. For example:

If Jack goes to bed late, then he will be tired.

Jack goes to bed late, but is not tired (or: Jack is tired, but he did not go to bed late).

Please write down as many factors as you can that could make this situation possible.

Depending on how many disablers and alternatives participants generate, the conditionals are subdivided in those having (1) many disablers and many alternatives, (2) many disablers and few alternatives, (3) few disablers and many alternatives, and (4) few disablers and few alternatives. These conditionals are then embedded in MP, MT, AC, and DA inferences and presented to a new group of participants who have to indicate how sure they are that the conclusion can be drawn. The second group of participants never sees the alternatives and disablers. Nevertheless, the acceptability of the conclusion is a function of the number of disablers and alternatives. Participants accept fewer MP and MT conclusions when conditionals have many instead of few disablers, and accept fewer AC and DA conclusions when conditionals have many instead of few alternatives (e.g., Cummins, 1995; De Neys, Schaeken, & d'Ydewalle, 2003a, 2003b; Gazzo Castañeda & Knauff, 2018).

2.3 The Importance of Content and Knowledge

One of the main reasons why human reasoning is defeasible stems from people's background knowledge and their prior beliefs about the content of conditionals (De Neys et al., 2003a, 2003b; Dieussaert, De Neys, & Schaeken, 2005; Evans & Over, 2004; Johnson-Laird & Byrne, 2002; Oaksford & Chater, 2007). Indeed, the relation between knowledge, beliefs, and content is not trivial (see chapter 13.3 by Knauff, this handbook). However, here we simply use the term "knowledge" in the sense of true or justified belief.

As shown by the findings with the covert paradigm, defeasibility does not require an explicit presentation of new evidence; instead, people activate their beliefs about the content of conditionals even if no explicit disablers or alternatives are presented (De Neys et al., 2003b; Vadeboncoeur & Markovits, 1999). Therefore, prior beliefs are an important prerequisite for defeasible reasoning. Without background knowledge (in the sense of true beliefs), people cannot know which information is a disabler or an alternative that can affect inferences. Experimentally, this has been shown in studies comparing participants' acceptance of conclusions for problems with unfamiliar and familiar content (Cummins, 1995; Gazzo Castañeda & Knauff, 2020; Markovits, 1986), by comparing the weighting of domain-specific disablers by laypeople and experts (Gazzo Castañeda & Knauff, 2016), and recently by varying the familiarity of the person or object in a conditional (Gazzo Castañeda & Knauff, 2019). In all these studies, the results are consistent: when participants do not have domain knowledge, they also make more MP, MT, AC, and/or DA inferences. In other words, they do not know the circumstances that could defeat the conclusions. It is therefore argued that the retrieval of disablers and alternatives affects the *perceived* sufficiency and necessity relations between *p* and *q*: disablers make the antecedent less sufficient for the consequent so that fewer MP and MT conclusions are accepted, and alternatives make the antecedent less necessary for the consequent so that fewer AC and DA conclusions are accepted (Cummins, 1995; Thompson, 1994, 1995).

Another factor related to background knowledge that is important for defeasible reasoning is the associative strength between the premises and the disablers in memory. Some disablers are more strongly associated with one's prior knowledge of how to *prevent q* from happening, and some alternatives are more strongly associated with one's prior knowledge of how to *cause q* (De Neys et al., 2003a; Quinn & Markovits, 1998). For instance, in our initial example "If Jack goes to bed late, then he will be tired," the disabler "Jack is able to sleep in" might be a stronger disabler than "Jack takes a cold morning shower." Therefore, both the number and the associative strength of disablers and alternatives affect people's acceptance of conclusions. The more of these strongly associated disablers or alternatives exist, the more probable it is that at least one will be retrieved, and the more probable it is that a conclusion will be withdrawn (De Neys et al., 2003a, 2003b; Vadeboncoeur & Markovits, 1999). Initially, an "all-or-nothing phenomenon" was assumed (De Neys et al., 2003b, p. 582): as soon as a person retrieves one disabling condition or alternative, the corresponding conclusion is rejected (cf. Vadeboncoeur & Markovits, 1999). However, in a series of experiments, De Neys and colleagues (2003b) showed that the consideration of disablers and alternatives is gradual: every additional disabler or alternative has an impact on the degree to which people accept a conclusion. Therefore, people's acceptance of the conclusion is related in a monotonic fashion to the number of known disablers and alternatives.

Besides knowledge about the number and strength of disablers and alternatives, also their frequency of occurrence is important. Geiger and Oberauer (2007)[1] noticed that it is important to distinguish between the number of different disablers and the frequency with which they are thought to occur (i.e., *p* & not-*q* cases). Sometimes, there are many disablers for one conditional, but their frequency of occurrence is low (and vice versa). For instance, for the conditional "If you open the fridge, then the light inside goes on," there are many disablers (e.g., broken fridge, broken light bulb, no electricity), but their overall occurrence is low (usually, the light goes on when the fridge is opened). When both factors are disentangled, the frequency of exceptions seems to predict people's withdrawal of conclusions more accurately than the number of disablers (Geiger & Oberauer, 2007). Defeasibility in human reasoning is thus not a phenomenon in which a conclusion is either accepted or rejected categorically. Instead, the withdrawal of conclusions happens by degrees and depends on how much one knows about the circumstances that may affect the conditional relationship.

2.4 The Importance of the Experimental Context

How the context influences defeasible reasoning has not received as much attention as content-related factors. However, contextual factors are nonetheless interesting as they can moderate the degree to which participants consider potential disablers and alternatives. For instance, Vadeboncoeur and Markovits (1999) showed that people's automatic activation of background knowledge about disablers can be inhibited by instructions emphasizing the logical necessity of conditional inferences. Furthermore, Markovits, Lortie Forgues, and Brunet (2010) showed that people's tendency to consider the frequency of exceptions can be inhibited by the presentation of a dichotomous response format. They showed that only when scaled response formats were available did information about the frequency of exceptions affect the acceptance of conclusions. But when there was only a dichotomous response format, then people rejected the conclusion already in light of a single disabler. Finally, the consideration of disablers and alternatives can also be moderated by the introduction of words questioning the believability of the conditional or the disabler. George (1997), for instance, conducted experiments where conditional statements were either presented traditionally as "if *p* then *q*" or with an additional "very probable" or "not very probable" before the conditional. He found that the certainty of the conditional statement influenced participants' belief in the conclusion. Stevenson and Over (1995) elicited similar effects by introducing doubt in the disabler. Participants were confronted with conditionals such as "If John goes fishing, he will have a fish supper" and disablers such as "If John catches a fish, he will have a fish supper." However, when participants were told, for example, that John is always lucky, they rejected fewer conclusions.

3. The Empirical Investigation of Belief Revision

In defeasible reasoning, people change their conclusions (derived beliefs) in light of new evidence. In belief revision, people instead have to make changes in their system of prior beliefs. This system of beliefs can be rather complex and interconnected, as we know from research in AI and philosophy. Psychologists are usually more pragmatic and often just treat the set of premises in an inference problem as the set of prior beliefs. Although this is a rather strong simplification, it is justified in many experimental settings in the psychological laboratory. Belief revision is then based on the assumption that a set of premises (beliefs) has to be consistent. Therefore, in our meditation example, Jack has to revise his beliefs: knowing that he did not fall asleep, he cannot believe that meditating makes him sleep. He has to find a way to reconcile his beliefs with what he observed. But which belief will he abandon? In AI and philosophy, two main accounts of which belief should be rejected have been proposed: minimal change (Harman, 1986) and epistemic entrenchment (Gärdenfors, 1988). According to the principle of *minimal change*, one should in case of inconsistencies make just enough changes in the belief set as are necessary to reach consistency. The principle of *epistemic entrenchment* arises from the idea that some beliefs are less entrenched than others and are therefore more easily given up in cases of conflict. Minimal change and epistemic entrenchment are well established in many formal theories of belief revision (e.g., Alchourrón, Gärdenfors, & Makinson, 1985; Gärdenfors, 1988; see chapter 5.2 by Rott and chapter 5.3 by Kern-Isberner, Skovgaard-Olsen, & Spohn, both in this handbook). However, from a psychological point of view, it is not clear how minimal change and epistemic entrenchment are implemented cognitively (cf. Elio & Pelletier, 1997). Therefore, many psychological studies have been carried out to understand the psychological processes behind these concepts and belief revision.

3.1 Experimental Findings

Psychological experiments on belief revision usually follow the same structure: participants are confronted with problems consisting of two (or more) premises and an additional fact. This additional fact, however, is inconsistent with the premises. Participants are then told that the additional fact is true but that the premises are uncertain. The task of the participants is to decide which of the two premises they believe less or abandon. One of the premises usually contains a general statement, such as a conditional (e.g., "If *p* then *q*") or a universally quantified statement (e.g., "All *A*s are *B*s" or "No *A*s are *B*s"). Let us illustrate this with an example from conditional reasoning:

If you meditate in the evening, then you fall asleep more easily.

You meditate in the evening.

But: You do not fall asleep more easily.

You will have noticed that the premises are the same as in MP. The additional fact, however, is not the conclusion that follows from the MP inference. It is even inconsistent with the set of premises. Participants thus have to indicate which premise they abandon: the conditional or the categorical one. In a similar way, belief revision can also be tested with MT:

If you meditate in the evening, then you fall asleep more easily.

You did not fall asleep more easily.

But: You meditated in the evening.

Again, participants have to decide whether they want to discard the conditional or the categorical premise.

The experimental findings on belief revision are heterogeneous. The first psychological studies on belief revision were conducted with quantified statements. Participants had to choose between revising premises with a universal quantifier (e.g., "All vertebrates have a backbone") or with particular facts (e.g., "This amoeba is not a vertebrate"). They consistently chose to revise the particular fact rather than the premise with the universal quantifier (e.g., Revlin, Cate, & Rouss, 2001; Revlis & Hayes, 1972; Revlis, Lipkin, & Hayes, 1971). The authors explained their findings by arguing that statements with universal quantifiers can be seen as general laws and are therefore more entrenched and believed more. However, years later, other studies had opposite results. In an influential study, Elio and Pelletier (1997) created conditional inference problems and found that the conditional premise—which is also a general law—is abandoned more readily than the categorical premise. This effect was more pronounced for MP than for MT. The authors argue that the conditional is revised more often because if–then "regularities" describe interdependences between particular facts and are therefore more uncertain (Elio & Pelletier, 1997). Elio and Pelletier's findings have been replicated several times (e.g., Byrne & Walsh, 2005, experiment 1; Elio, 1997, experiment 1; Politzer & Carles, 2001).

One explanation for these conflicting findings is the premises' content. Studies finding a preference for believing the conditional or universal premise used content-rich material, for which the reasoner has background knowledge. By contrast, studies finding a preference for *disbelieving* the conditional or universal premise used unfamiliar content, for which the reasoner does not have prior knowledge. This was tested by Byrne and Walsh (2005). In two experiments, they varied the reasoner's familiarity with the conditional premise by creating conditionals for which the reasoner had no knowledge ("If the ruin was inhabited by Pings it had a forcefield surrounding it") or had knowledge ("If water was poured on the campfire the fire went out"). After presenting to the participants the unquestionable fact, they had to write down which premise they revised. When the content was unfamiliar, participants preferred to revise the conditional premise. But when the content was familiar, participants preferred to revise the categorical premise. Similarly, in another study, Politzer and Carles (2001) varied the plausibility of conditionals and found that more plausible conditionals were revised less often than less plausible ones. Correspondingly, Wolf, Rieger, and Knauff (2012) also found that conditional premises uttered by less trustworthy sources are revised more often than those uttered by more trustworthy ones. Along these lines, the number of disablers affects belief revision too. Elio (1997) constructed belief revision problems with conditionals that differed in their number of disablers (including the ones from Cummins et al., 1991). She found that the conditional premise was believed less, and abandoned more often, for conditionals with many disablers compared to conditionals with few disablers.

Overall, these studies show that human belief revision depends on the content of the premises and the trustworthiness of the information source. Revlis's assumption that premises containing general rules, such as conditionals, are more entrenched because they are lawlike applies only as long as this general rule is known by, and plausible for, the reasoner. When the conditional or universal premise is unfamiliar instead, then reasoners have no problems in revising such beliefs. In other words, entrenchment depends on the knowledge base of the reasoner (Byrne & Walsh, 2005; Politzer & Carles, 2001). However, this content-dependent approach cannot explain all revision choices. For instance, Elio and Pelletier (1997) found that the preference for revising the conditional premise instead of the categorical one was more pronounced for MP than for MT. This effect has also been replicated several times (e.g., Khemlani & Johnson-Laird, 2011; Revlin et al., 2001). There are thus additional, inference-dependent processes involved in belief revision. We present some explanations for this MP–MT asymmetry in the next section.

3.2 Spatial Belief Revision

The empirical findings on belief revision we have presented so far were concerned with conditionals. This is understandable because "if–then" statements are closely related to our beliefs about the world. Moreover, the experimental paradigm follows the philosophical tradition in which an inconsistency arises between an additional fact and the conclusion that can be derived from a set of premises. But in everyday life, people are often forced to change other kinds of beliefs. For instance, when following directions or trying to orient themselves in foreign cities, people also have to change beliefs about spatial location when they are lost.

Our group was the first that explored belief revision processes in the domain of spatial reasoning. In spatial reasoning, people infer new spatial relations between objects in space from information that is already given (Knauff, 2013; Ragni & Knauff, 2013). But what happens if new information disagrees with the prior information about the objects' locations in space? In our belief revision paradigm, participants get two or more premises describing the spatial relations between a set of objects. Afterward, an additional fact is presented that conflicts with the spatial arrangement consistent with the premises. Participants have to decide which objects they relocate in order to reach consistency. For example:

A is to the left of *B*.

B is to the left of *C*.

But: *A* is to the right of *C*.

The two premises allow the construction of the linear spatial arrangement *A*–*B*–*C*. However, the contradictory fact says that *A* is to the right of *C*. There are two ways to solve this inconsistency: either moving *A* to the right of *C* or moving *C* to the left of *A*. Following the psychological and linguistic tradition on the functional asymmetry between arguments, *A* is called the *to-be-located object* (LO) and *C* the *reference object* (RO) (Landau & Jackendoff, 1993). The empirical findings show that participants prefer to relocate the LO instead of the RO (Knauff, Bucher, Krumnack, & Nejasmic, 2013), although this effect can be modulated by properties of the objects such as size and movability (Nejasmic, Bucher, & Knauff, 2015).

4. Psychological Theories of Defeasible Reasoning and Belief Revision

So far, we have described how defeasible reasoning and belief revision are investigated in psychology and given an overview of the main experimental findings. Now we want to give an overview of the current psychological theories of defeasible reasoning and belief revision. This will, hopefully, facilitate the understanding of the relatively complex (and not always consistent) pattern of effects and provide insights into the cognitive processes behind defeasible reasoning and belief revision.

4.1 The Theory of Mental Models

The theory of mental models (Johnson-Laird, 2010; Johnson-Laird & Byrne, 1991; see chapter 2.3 by Johnson-Laird, in this handbook) describes human reasoning as a semantic process where people use the meaning of premises to construct internal mental representations—the models—of what would be the case if the premises were true. These mental models are inspected and evaluated in order to find information that is not already given in the premises. The construction of mental models follows the principle of truth, meaning that only what is true is represented. Therefore, following the postulates of classical logic, a conditional "If *p* then *q*" allows the construction of the following mental models:

p	*q*
not-*p*	*q*
not-*p*	not-*q*

A conclusion is considered *true* if it follows necessarily from the premises. A conclusion is considered *false* if there is a counterexample to it, that is, a model in which the premises are true but the conclusion is false. For instance, in the models above, we can see that if *p* is true, *q* is also true. There is no model in which *p* is true but *q* is false. Therefore, the MP conclusion is true. Along the same lines, also the validity of MT and the invalidity of AC and DA can be shown.

Furthermore, the model theory considers the restrictions from working memory. It is postulated that people do not consider all mental models from the beginning. Instead, they only construct one mental model, the *explicit* mental model *p* & *q*, which represents the information explicitly provided in the conditional:

p *q*
. . .

All other possible models (not-*p* and *q*, and not-*p* and not-*q*) are only represented implicitly, denoted by the three dots. Only if required by the task are these implicit models fleshed out into fully explicit models. Several experiments have shown that people prefer to construct specific models (*preferred mental models*) while ignoring other possibilities that are also consistent with the premises (*neglected mental models*) (Jahn, Knauff, & Johnson-Laird, 2007; Knauff, 2013). Moreover, the sequence of models generated is not random but follows a specific, cognitively efficient model variation mechanism (Ragni

& Knauff, 2013). This theory has recently been fundamentally revised and extended to account for ordinary people's reasoning beyond deduction (Hinterecker, Knauff, & Johnson-Laird, 2016; Johnson-Laird, Khemlani, & Goodwin, 2015b).

Defeasible reasoning Mental model theory accounts for defeasible reasoning by assuming that reasoners introduce disablers as additional antecedents (Byrne et al., 1999; Johnson-Laird & Byrne, 2002). Imagine, for instance, that a reasoner is confronted with the conditional "If Jack goes to bed late, then he will be tired" and considers the disabler that he can sleep in. This results in the following mental model:

bed late	sleeping in	tired
	. . .	

Because reasoners know that after sleeping in, people are usually not tired, they will additionally construct the following models, based on general knowledge:

sleeping in	not tired
not sleeping in	tired
not sleeping in	not tired

Thus, when people have to decide whether *q* follows from *p*, they combine the models from background knowledge with the models of the premises. However, the combination of these models would yield a contradiction:

bed late	sleeping in	tired	not tired

Therefore, Johnson-Laird and Byrne (2002) propose the principle of *pragmatic modulation*: in such cases, the model from prior knowledge has priority, allowing reasoners to conclude that, if Jack goes to bed late but is able to sleep in, he will not be tired. The model theory thus describes the consideration of disablers in terms of counterexample search in general knowledge. The impact of the frequency of exceptions is accounted for by assuming that the probability of an event can be inferred from the proportion of models in which the event occurs (e.g., Johnson-Laird, 2001; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999). For instance, when reasoning about the conditional "If Jack goes to bed late, then he is tired," reasoners may construct the following mental models, containing the chances that *q* actually follows from *p*:

bed late	tired	3
bed late	not tired	1
not bed late	tired	2
not bed late	not tired	4

In this case, there are three chances that Jack is tired out of four chances that he goes to bed late (see also Johnson-Laird, Khemlani, & Goodwin, 2015a).

Belief revision According to mental model theory, belief revision happens in three steps (Johnson-Laird, Girotto, & Legrenzi, 2004; Legrenzi & Johnson-Laird, 2005): first, participants try to construct a mental model where all premises are satisfied. If this is not possible, participants infer that the set of premises is inconsistent. Second, participants try to solve the inconsistency. If the contradicting fact only conflicts with a single premise, then this premise is disbelieved. However, if the contradicting fact conflicts with more premises, then people will revise the premise whose model is in conflict with the contradictory fact or that fails to represent the fact. This process is called the *mismatch principle* (Johnson-Laird et al., 2004). For instance, in the case of MP, people are confronted with the conditional "If *p* then *q*," the categorical premise *p*, and the contradictory fact not-*q*. This contradictory fact is clearly in conflict with the explicit model from the conditional premise (*p q*). Therefore, the mismatch principle predicts that, for MP, the conditional premise will be abandoned. In MT, by contrast, people are confronted with the conditional "If *p* then *q*," the categorical premise not-*q*, and the contradictory fact *p*. In this case, *p* is not in conflict with the explicit model and thus the categorical premise will be abandoned instead of the conditional. This explains why the conditional premise is more often disbelieved in MP than MT. Finally, after belief revision, participants search for a *diagnostic explanation* that resolves the inconsistency they observed. For instance, in our example about the failed meditation, such an explanation could be "It was the wrong meditation and therefore he did not fall asleep" (see Legrenzi & Johnson-Laird, 2005).

4.2 Probabilistic Theories

Probabilistic theories assume Bayesian probability theory as the right norm for human reasoning (Evans, 2012; Oaksford & Chater, 2007; see chapter 4.5 by Chater & Oaksford and chapter 4.6 by Oberauer & Pessach, both in this handbook). Proponents of probabilistic theories of reasoning assume that people treat conditionals probabilistically by understanding the probability of a conditional, $P(\text{if } p \text{ then } q)$, as the conditional probability of *q* given *p*, $P(q \mid p)$—a relationship known as "the Equation" (Edgington, 1995; Evans & Over, 2004). The conditional probability can be calculated by performing the Ramsey test (Ramsey, 1929/1990; see also Evans, Handley, & Over, 2003). According to the Ramsey test, people first suppose that *p* holds. On the basis of this hypothetical belief, people then estimate the probability of *q*. For instance, in the case of "If Jack goes to bed late, then he will be tired," reasoners first assume that Jack goes to bed late. Then they start thinking about how probable it is that a person actually feels tired after going to bed late ($P(p \& q)$) and how probable it is that this is not the case ($P(p \& \text{not-}q)$). The higher $P(p \& q)$ is relative to $P(p \& \text{not-}q)$, the higher is the conditional probability $P(q \mid p)$ (Evans & Over, 2004; Evans et al., 2003). And the higher this conditional probability, the more probable it is that an MP conclusion is accepted (Oaksford, Chater, & Larkin, 2000).

Defeasible reasoning The notion of conditional probabilities captures defeasible reasoning directly. In fact, nonmonotonic and defeasible reasoning were the purposes probabilistic theories were developed for in the first place (Oaksford & Chater, 2013). All information that influences the conditional probability affects people's acceptance of conclusions. Therefore, disablers affect reasoning: disablers lower the conditional probability $P(q \mid p)$, and as a consequence, also fewer MP and MT conclusions are drawn (Weidenfeld, Oberauer, & Hörnig, 2005). In a similar way, alternatives affect inferences too: alternatives lower the conditional probability $P(p \mid q)$, and therefore fewer AC and DA conclusions are drawn (Verschueren, Schaeken, & d'Ydewalle, 2005). Further, given that probabilistic theories do not assume mental representations of concrete *p*-, *q*-, not-*p*-, or not-*q*-cases, they can explain why in addition to the number of disablers, also the overall frequency of exceptions affects inferences. Probabilistic approaches do not require people to construct concrete representations from general knowledge to be aware that *p* can happen without *q*. Instead, the effect of disablers is operationalized by the probability $P(p \& \text{not-}q)$, which is directly related to the frequency of exceptions.

Belief revision On the normative side, there are many accounts that explain belief change, for instance, through Bayesian conditionalization. However, from a psychological point of view, these accounts may be problematic (see Oaksford & Chater, 2013). Therefore, further research is needed to understand how people change their probability distributions after learning new information. For instance, Wolf, Rieger, and Knauff (2012) provided an account of how probabilities could describe the asymmetry in belief revision for MP and MT. The main assumption is that the contradictory fact changes the probabilities of the premises. For instance, in the case of belief revision after MP, the fact not-*q* is presented as true; therefore, $P(\text{not-}q) = 1$. This implies that $P(q) = 0$, and consequently, $P(q \mid p) = P(\text{if } p \text{ then } q) = 0$. However, in the case of belief revision after MT, the fact *p* is presented as true; therefore, $P(p) = 1$. Consequently, $P(q \mid p) = P(\text{if } p \text{ then } q) = q$. The comparison of these probabilities shows why the conditional premise is often believed less for MP than MT.

The influence of plausibility and trustworthiness on belief revision is described by probabilistic theories through the role of long-term memory. The main idea is that long-term memory about the content of a conditionals affects the corresponding conditional probability (Oaksford & Chater, 2013). Similarly, also knowledge about plausibility and trustworthiness is part of our knowledge system. Therefore, participants give lower probabilities to implausible conditionals and to conditionals uttered by untrustworthy sources. As a consequence, these are believed less and abandoned more readily (cf. Wolf et al., 2012).

4.3 Dual-Process Theories

Dual-process theories propose that people use *two kinds of thinking* or *modes of processing* that are employed depending on factors such as tasks, time constraints, or cognitive capacity (Evans, 2008; Evans & Stanovich, 2013; see chapter 2.4 by Fiedler, Prager, & McCaughey, this handbook). People can reason heuristically, fast, and effortlessly (known as "System 1" or "Type 1" processes), and they can reason analytically, slowly, and dependently on working memory (known as "System 2" or "Type 2" processes) (e.g., Evans, 2008; Kahneman & Frederick, 2005). It is argued that System 1 serves as a default system, whose responses can be either overridden or supported by System 2 (Evans & Stanovich, 2013).

Defeasible reasoning As we have shown, disablers and alternatives can affect reasoning in two ways: either as single instances or through the frequency of *p* & not-*q*- and not-*p* & *q*-cases. Verschueren et al. (2005) argue that both aspects are important and can be unified in a dual-process approach. They argue that the consideration of single disablers and alternatives is an analytic process—as proposed by mental model theory. In contrast, the consideration of the frequency information is considered a heuristic process, in which the overall likelihood of *q* given *p* and of *p* given *q* matters—as proposed by the probabilistic theories. Verschueren et al.'s (2005) dual-process model has been tested experimentally and corroborated several times (e.g., Markovits, Brisson, & de Chantal, 2015, 2017). As predicted by dual-process theories, fast responses are best explained by the likelihood of *q* given *p* (for MP and MT inferences) or the likelihood of *p* given *q* (for AC and DA inferences), slow responses are best described by the availability of disablers or alternatives, and conclusions based on likelihoods can be overridden by specific disablers and alternatives.

Belief revision A dual-process approach to belief revision can be found in Wolf and Knauff (2008). In their experiments, they varied the probability of the conditional premise (0%, 50%, or 100%). They found that when the conditional premise was highly probable or highly improbable, participants used this probability information as a heuristic for belief revision: they believed the conditional premise more for highly probable conditionals but believed the categorical premise more for highly improbable conditionals. However, when the probability of the conditional was around 50%, people reasoned more analytically, and the mismatch principle from mental model theory predicted belief revision choices: the conditional premise was believed more for MT than for MP.

5. Defeasible Reasoning or Belief Revision?

So far, we have treated defeasible reasoning and belief revision separately. The main reason is that both functions are often discussed within different theoretical (normative) frameworks and communities. In defeasible reasoning, conclusions are withdrawn in light of new evidence, while in belief revision, the new evidence makes people abandon previous beliefs (in the psychology lab typically operationalized as premises). However, a germane question is whether defeasible reasoning and belief revision do indeed rely on distinct *cognitive* processes—a crucial descriptive question from a psychological point of view. In fact, the psychological literature on defeasible reasoning and belief revision shows that both depend on the content of the premises. In defeasible reasoning, many disablers and unbelievable content increase people's willingness to reject conclusions from conditional premises. In a similar way, in belief revision, conditionals with many disablers and implausible content are revised more often than those with fewer disablers or plausible content. These similarities were noticed by Dieussaert et al. (2005). They conducted a belief revision experiment using conditionals with many and with few disablers and asked the participants twice about their belief in the conditional premise: before presenting the contradicting fact and after. They found no significant differences in the believability ratings before and after the introduction of the inconsistency. This shows that people's disbelief in the conditional premise was not the outcome of a process of belief revision but a product of people's consideration of disablers. Dieussaert and colleagues therefore argue that people's withdrawal of conclusions in conditional reasoning and in belief revision are "two sides of the same coin" (Dieussaert et al., 2005, p. 29; see also Gärdenfors, 1990).

At first sight, the findings by Dieussaert et al. (2005) seem counterintuitive. In belief revision, participants abandon premises in the search for consistency, while in defeasible reasoning, there is no inconsistency at all. Nevertheless, there is evidence suggesting that this distinction is only a result of differing experimental paradigms. For instance, Walsh and Johnson-Laird (2009) showed that when participants are asked to explain inconsistencies, they rarely just negate the conditional premise but rather explain inconsistencies by naming disablers that question the causal link between antecedent and consequent (see also Johnson-Laird et al., 2004). In a later experiment, Khemlani and Johnson-Laird (2011) showed that in case of inconsistencies, participants prefer to give *explanations* instead of direct refutations of the premises. That is, participants seem to resolve inconsistencies by finding disablers, without the need of a revision in the strict sense. Similarly, also Politzer and Carles (2001) found that when participants can choose between completely abandoning one of the premises and only casting doubt on them, they prefer to cast doubt. These findings suggest that under everyday circumstances, people do not spontaneously abandon existing beliefs but instead try to find explanations for inconsistencies in the set of beliefs. This has consequences for the relation between defeasible reasoning and belief revision. An example:

If you meditate in the evening, then you fall asleep more easily.

You meditate in the evening.

But: You do not fall asleep more easily.

In a belief revision paradigm, participants have to choose which premise they reject. However, in everyday life, participants would rather try to explain the inconsistency by the introduction of a disabler:

If you meditate in the evening, then you fall asleep more easily.

If it is the wrong meditation, then you do not fall asleep more easily.

You meditate in the evening.

You do not fall asleep more easily.

Now the inference resembles one of defeasible reasoning: by introducing the disabler, the formally inconsistent fact turns into the "conclusion" that participants would have accepted in a defeasible reasoning paradigm. The search for explanations plays a vital role in these cognitive processes, albeit one that is not so frequently considered in normative accounts of defeasible reasoning and belief revision.

6. Concluding Remarks and Further Perspectives

Many psychological studies help us to understand human defeasible reasoning and belief revision. Still, there is need for further research. For instance, the principle of truth of mental model theory has difficulties in capturing people's consideration of *p* & not-*q*-cases during reasoning (Geiger & Oberauer, 2010). Similarly, probabilistic theories are only computational theories (according to Marr's, 1982, terminology) that do not describe the cognitive processes behind reasoning (Oaksford & Chater, 2003). However, cognitive theories of reasoning must explain not only *what* is computed but also *how* it is computed in the human mind (Hinterecker, Knauff, & Johnson-Laird, 2019; see the introductory chapter by Knauff & Spohn, this handbook). Dual-process theories have been criticized for not describing qualitative differences but only quantitative differences in reasoning (Osman, 2013) and for just exemplifying "the backwards development from precise theories to surrogates" (Gigerenzer, 2011, p. 739; but see chapter 1.2 by Evans, this handbook).

What remains clear from our review is that still more research is needed to understand the cognitive processes that underlie human defeasible reasoning and belief revision. Here, more collaboration between psychology, philosophy, and AI could be helpful. There are many normative and formal theories on defeasible reasoning (e.g., Kraus, Lehmann, & Magidor, 1990; Spohn, 2012; see chapter 5.3 by Kern-Isberner, Skovgaard-Olsen, & Spohn, this handbook) and belief revision (e.g., Alchourrón et al., 1985; see chapter 5.2 by Rott, in this handbook), which can be helpful in cognitive psychology, too. At the same time, however, these formal theories should also consider psychological findings and adapt their formalisms to the actual cognition and behavior of real people. We are well aware of the different views on this *is–ought problem* (see Knauff and Spohn's introduction to this handbook). Yet, it is obvious that there are still too few links between the normative and the descriptive work on defeasibility and belief revision. On the one hand, many of the formalisms from philosophy and AI have not been tested regarding their capability to model actual human reasoning (cf. Ragni, Eichhorn, Bock, Kern-Isberner, & Ping Ping Tse, 2017). On the other hand, psychological research has neglected nonmonotonicity for a long time. Therefore, psychologists have often prematurely considered people irrational, just because they did not follow the norms of classical logic. In fact, many of the psychological studies we have reported here do not even use the technical term "defeasible reasoning," although they have investigated the circumstances under which it is rational to withdraw otherwise valid conclusions. To avoid such dangers of disciplinary division, we need a more interdisciplinary approach. Ideally, this research combines the strengths of psychology, philosophy, and AI to obtain a broader understanding of how a cognitive system, be it real or abstract, human or machine, can flexibly respond to new information and therefore change its belief system in a rational way.

Note

1. Geiger and Oberauer (2007) only investigated the relation between the number of disablers and the frequency of exceptions. However, it is also conceivable that the frequency of alternatives affects inferences.

References

Alchourrón, C. E., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, *50*, 510–530.

Antoniou, G. (1997). *Nonmonotonic reasoning*. Cambridge, MA: MIT Press.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*, 61–83.

Byrne, R. M. J., Espino, O., & Santamaria, C. (1999). Counterexamples and the suppression of inferences. *Journal of Memory and Language*, *40*, 347–373.

Byrne, R. M. J., & Walsh, C. R. (2005). Resolving contradictions. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi* (pp. 117–122). Hove, England: Psychology Press.

Cummins, D. D. (1995). Naive theories and causal deduction. *Memory & Cognition*, *23*, 646–658.

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*, 274–282.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003a). Causal conditional reasoning and strength of association: The disabling condition case. *European Journal of Cognitive Psychology*, *15*, 161–176.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003b). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, *31*, 581–595.

Dieussaert, K., De Neys, W., & Schaeken, W. (2005). Suppression and belief revision, two sides of the same coin? *Psychologica Belgica*, *45*, 29–46.

Edgington, D. (1995). On conditionals. *Mind*, *104*, 235–329.

Elio, R. (1997). What to believe when inferences are contradicted: The impact of knowledge type and inference rule. In M. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (pp. 211–216). Mahwah, NJ: Erlbaum.

Elio, R., & Pelletier, F. J. (1997). Belief change as propositional update. *Cognitive Science*, *21*, 419–460.

Evans, J. St. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278.

Evans, J. St. B. T. (2012). Questions and challenges for the new psychology of reasoning. *Thinking & Reasoning*, *18*, 5–31.

Evans, J. St. B. T., Handley, S. J., & Over, D. E. (2003). Conditionals and conditional probability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 321–355.

Evans, J. St. B. T., Newstead, S. E., & Byrne, R. (1993). *Human reasoning: The psychology of deduction*. Hove, England: Erlbaum.

Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford, England: Oxford University Press.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*, 223–241.

Gärdenfors, P. (1988). *Knowledge in flux*. Cambridge, MA: MIT Press.

Gärdenfors, P. (1990). Belief revision and nonmonotonic logic: Two sides of the same coin? In L. Carlucci Aiello (Ed.), *Proceedings of the Ninth European Conference on Artificial Intelligence* (pp. 768–773). London, England: Pitman Publishing.

Gazzo Castañeda, L. E., & Knauff, M. (2016). Defeasible reasoning with legal conditionals. *Memory & Cognition*, *44*, 499–517.

Gazzo Castañeda, L. E., & Knauff, M. (2018). Quantifying disablers in reasoning with universal and existential rules. *Thinking & Reasoning*, *24*, 344–365.

Gazzo Castañeda, L. E., & Knauff, M. (2019). The specificity of terms affects conditional reasoning. *Thinking & Reasoning*, *25*, 72–93.

Gazzo Castañeda, L. E., & Knauff, M. (2020). Everyday reasoning with unfamiliar conditionals. *Thinking & Reasoning*. doi:10.1080/13546783.2020.1823478

Geiger, S. M., & Oberauer, K. (2007). Reasoning with conditionals: Does every counterexample count? It's frequency that counts. *Memory & Cognition*, *35*, 2060–2074.

Geiger, S. M., & Oberauer, K. (2010). Towards a reconciliation of mental model theory and probabilistic theories of conditionals. In M. Oaksford & N. Chater (Eds.), *Cognition and conditionals: Probability and logic in human thinking* (pp. 289–307). Oxford, England: Oxford University Press.

George, C. (1997). Reasoning from uncertain premises. *Thinking & Reasoning*, *3*, 161–189.

Gigerenzer, G. (2011). Personal reflections on theory and psychology. *Theory & Psychology*, *20*, 733–743.

Harman, G. (1986). *Change in view*. Cambridge, MA: MIT Press.

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2016). Modality, probability, and mental models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1606–1620.

Hinterecker, T., Knauff, M., & Johnson-Laird, P. N. (2019). How to infer possibilities: A reply to Oaksford et al. (2018). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(2), 298–301.

Jahn, G., Knauff, M., & Johnson-Laird, P. N. (2007). Preferred mental models in reasoning about spatial relations. *Memory & Cognition*, *35*, 2075–2087.

Johnson-Laird, P. N. (2001). Mental models and deduction. *Trends in Cognitive Sciences*, *5*, 434–442.

Johnson-Laird, P. N. (2010). Mental models and human reasoning. *Proceedings of the National Academy of Sciences*, *107*, 18243–18250.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove, England: Erlbaum.

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*, 636–678.

Johnson-Laird, P. N., Girotto, V., & Legrenzi, P. (2004). Reasoning from inconsistency to consistency. *Psychological Review*, *111*, 640–661.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015a). Logic, probability, and human reasoning. *Trends in Cognitive Science*, *19*, 201–214.

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015b). Response to Baratgin et al.: Mental models integrate probability and deduction. *Trends in Cognitive Sciences*, *19*, 548–549.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review*, *106*, 62–88.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–294). Cambridge, England: Cambridge University Press.

Khemlani, S., & Johnson-Laird, P. N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, *64*, 2276–2288.

Knauff, M. (2013). *Space to reason: A spatial theory of human thought*. Cambridge, MA: MIT Press.

Knauff, M., Bucher, L., Krumnack, A., & Nejasmic, J. (2013). Spatial belief revision. *Journal of Cognitive Psychology*, *25*, 147–156.

Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, *44*, 167–207.

Landau, B., & Jackendoff, R. (1993). “What” and “where” in spatial language and spatial cognition. *Behavioral and Brain Sciences*, *16*, 217–265.

Legrenzi, P., & Johnson-Laird, P. N. (2005). The evaluation of diagnostic explanations for inconsistencies. *Psychologica Belgica*, *45*, 19–28.

Markovits, H. (1986). Familiarity effects on conditional reasoning. *Journal of Educational Psychology*, *78*, 492–494.

Markovits, H., Brisson, J., & de Chantal, P.-L. (2015). Additional evidence for a dual-strategy model of reasoning: Probabilistic reasoning is more invariant than reasoning about logical validity. *Memory & Cognition*, *43*, 1208–1215.

Markovits, H., Brisson, J., & de Chantal, P.-L. (2017). Logical reasoning versus information processing in the dual-strategy model of reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 72–80.

Markovits, H., Lortie Forgues, H., & Brunet, M.-L. (2010). Conditional reasoning, frequency of counterexamples, and the effect of response modality. *Memory & Cognition*, *38*, 485–492.

Marr, D. (1982). *Vision*. San Francisco, CA: Freeman.

Nejasmic, J., Bucher, L., & Knauff, M. (2015). Grounded spatial belief revision. *Acta Psychologica*, *157*, 144–154.

Oaksford, M., & Chater, N. (2003). Conditional probability and the cognitive science of conditional reasoning. *Mind & Language*, *18*, 359–379.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford, England: Oxford University Press.

Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning*, *19*, 346–379.

Oaksford, M., Chater, N., & Larkin, J. (2000). Probabilities and polarity biases in conditional inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 883–899.

Osman, M. (2013). A case study: Dual-process theories of higher cognition—Commentary on Evans & Stanovich (2013). *Perspectives on Psychological Science*, *8*, 248–252.

Politzer, G., & Bonnefon, J.-F. (2006). Two varieties of conditionals and two kinds of defeaters help reveal two fundamental types of reasoning. *Mind & Language*, *21*, 484–503.

Politzer, G., & Carles, L. (2001). Belief revision and uncertain reasoning. *Thinking & Reasoning*, *7*, 217–234.

Quinn, S., & Markovits, H. (1998). Conditional reasoning, causality and the structure of semantic memory: Strength of association as a predictive factor for content effects. *Cognition*, *68*, B93–B101.

Ragni, M., Eichhorn, C., Bock, T., Kern-Isberner, G., & Ping Ping Tse, A. (2017). Formal nonmonotonic theories and properties of human defeasible reasoning. *Minds and Machines*, *27*, 79–117.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*, 561–588.

Ramsey, F. P. (1990). General propositions and causality (original manuscript, 1929). In D. H. Mellor (Ed.), *Philosophical papers* (pp. 145–163). Cambridge, England: Cambridge University Press.

Revlin, R., Cate, C. L., & Rouss, T. S. (2001). Reasoning counterfactually: Combining and rending. *Memory & Cognition*, *29*, 1196–1208.

Revlis, R., & Hayes, J. R. (1972). The primacy of generalities in hypothetical reasoning. *Cognitive Psychology*, *3*, 268–290.

Revlis, R., Lipkin, S. G., & Hayes, J. R. (1971). The importance of universal quantifiers in a hypothetical reasoning task. *Journal of Verbal Learning and Verbal Behaviour*, *10*, 86–91.

Rumain, B., Connell, J., & Braine, M. D. (1983). Conversational comprehension processes are responsible for reasoning fallacies in children as well as adults: *If* is not the biconditional. *Developmental Psychology*, *19*, 471–481.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford, England: Oxford University Press.

Stevenson, R. J., & Over, D. E. (1995). Deduction from uncertain premises. *Quarterly Journal of Experimental Psychology*, *48A*, 613–643.

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*, 742–758.

Thompson, V. A. (1995). Conditional reasoning: The necessary and sufficient conditions. *Canadian Journal of Experimental Psychology*, *49*, 1–58.

Vadeboncoeur, I., & Markovits, H. (1999). The effect of instructions and information retrieval on accepting the premises in a conditional reasoning task. *Thinking & Reasoning*, *5*, 97–113.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). A dual-process specification of causal conditional reasoning. *Thinking & Reasoning*, *11*, 239–278.

Walsh, C. R., & Johnson-Laird, P. N. (2009). Changing your mind. *Memory & Cognition*, *37*, 624–631.

Weidenfeld, A., Oberauer, K., & Hörnig, R. (2005). Causal and noncausal conditionals: An integrated model of interpretation and reasoning. *Quarterly Journal of Experimental Psychology*, *58A*, 1479–1513.

Wolf, A. G., & Knauff, M. (2008). The strategy behind belief revision: A matter of judging probability or the use of mental models? In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 831–836). Austin, TX: Cognitive Science Society.

Wolf, A. G., Rieger, S., & Knauff, M. (2012). The effects of source trustworthiness and inference type on human belief revision. *Thinking & Reasoning*, *18*, 417–440.

The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

The MIT Press

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455