8

DUAL FRAMES IN CAUSAL REASONING AND OTHER TYPES OF THINKING

Masasi Hattori, David E. Over, Ikuko Hattori, Tatsuji Takahashi and Jean Baratgin

Introduction

Manktelow (2012, p. 89) discusses an intriguing poster he once saw on a local bus that made a causal claim. It read, 'FACT: Cigarettes and alcohol are the biggest cause of death by fire in the home.' He rightly remarks that it is not easy to pin down this claim. Causal reasoning is indispensable to living beings: we would soon die without it. But we will argue that, to be clear about causation, we must make a distinction between two types of causal induction. We can try to identify quickly and efficiently the causal, or at least relevant, factors for preventing death in our homes. Looking at some data, local health officials might swiftly detect that cigarette smoking and alcohol are such factors for death in the home and produce a poster with that information on it. But other, more scientific researchers might want deeper knowledge about how to predict and control a type of event, like death in the home, and would aim to distinguish between correlation and causation. With deeper research, they might, for example, be able to set limits to alcohol consumption, above which the danger greatly increases that it will cause people to make serious mistakes or misjudgements.

The primary claim of this chapter is that people's thinking, including causal reasoning, has two distinct processes, or ways of seeing things in the world, which we call *frames*. Among many factors that affect ways of recognising affairs in the world, causality is primary for framing, and we are starting with causal induction. People employ two types of processes for two types of causal induction: fast screening and the precise identification of a cause. People switch between two different 'perspectives', each corresponding to one of these two processes. We have labelled the first perspective as *A-frame* (A stands for attentional) and the second as *B-frame* (B stands for balanced). See Table 8.1. When perceiving causality through the A-frame, people are trying to detect a correlation between

TABLE 8.1 Characteristics of the two frames in causal reasoning

	A-frame	*B-frame*
Epistemic aim	Fast screening	Control
Thought style	Relevance mode	Differentiation mode
Focalisation	Positivity focus	Comparative view
Psychological symmetry	Asymmetrical	Symmetrical
Negation	Explicit (*X* vs. *not-X*)	Implicit (*X* vs. *Y*)
Cognitive process	Heuristic	Analytic
D-cell	Disregard	Respect
Base rate	Rare	Not rare
Causality scaling	Monopolar (null/effective)	Bipolar (preventive/generative)
Invasiveness	Observation	Intervention
Activeness	Passive	Active
Commitment	Uncommitted	Committed

a potential cause (hereafter *C*) and a target effect (hereafter *E*); this is the *relevance mode*. On the other hand, the B-frame facilitates a comparative examination between cases in which *C* exists and cases in which it does not, in order to detect how the occurrence of *E* is affected by the control of *C*; this is the *differentiation mode*.

Most studies of causal induction have assumed that there is just one type of induction (or learning). This includes normative rule-based theories (e.g. Cheng, 1997; White, 2003), associative accounts (e.g. Shanks, Lopez, Darby & Dickinson, 1996; Wasserman, Kao, Van Hamme, Katagiri, & Young, 1996) and Bayesian approaches (e.g. Griffiths & Tenenbaum, 2005, 2009; Lu, Liljeholm, Cheng & Holyoak, 2008). Aside from a few exceptional cases (e.g. Hattori & Oaksford, 2007), existing research has not shown a clear awareness of the difference between the two types of causal reasoning, but we will explore the distinction here. If one wants to avoid an apparent present danger or simply predict the near future, then it is not necessary to go so far as to know the causal structure. Rather, the important thing in such a case is to identify as quickly as possible the factors related to the problematic element, for example death in the home. Whether a relevant factor is a genuine cause is not a pressing issue; the important thing is to discover the most relevant factors from a small sample. The goal of the second type of causal induction is to control the effect, for example prevent accidents from alcohol consumption. In this case, it is important to gain an understanding of the causal structure. Causality usually has many factors, so it is necessary to clarify which ones have a direct causal relationship with each other, what the direction of causal relationship is and how strong the relationship is. Previous causal induction studies have dealt with normative models based on the B-frame and have not recognised the importance of causal reasoning in the A-frame. However, the A-frame is not only relevant to causal induction; it is also deeply related to errors and biases that have been clarified in research in fields other than causal induction. This assertion represents the secondary claim of this chapter.

A-frame: diagnostic probability, *D*-cell disregard and monopolar causality

The *dual-factor heuristic* (DFH) model proposed by Hattori and Oaksford (2007) represents A-frame causal induction in the simplest possible way:

$$H = \sqrt{P(E \mid C)P(C \mid E)} = \frac{a}{\sqrt{(a+b)(a+c)}}$$

In this model, a, b, c and d represent the frequencies for combinations of occurrence/non-occurrence of C and occurrence/non-occurrence of E as defined in Table 8.2. This model is notable for its incorporation of $P(C \mid E)$ and for the fact that the d-cell is not included as a variable. The latter feature will be discussed in a later section, and we now address the former feature.

Suppose people get food poisoning at a restaurant. The first step someone would take to find out the cause of the incident would likely be to ask the victims what they ate, in order to identify the problematic food item. Such an action is nothing other than the identification of a C with high $P(C \mid E)$. If E is rare (i.e. if a *rarity assumption*, Oaksford & Chater, 1994, is established), then the $P(C \mid E)$ could be extremely useful data for screening relevant factors. In this context, a rarity assumption means that the occurrence frequency is by default assumed relatively small. For example, in a case such as 'if you turn the ignition key (C), the car will start (E),' we usually consider both events – turning the ignition key and the car starting – as having low occurrence frequencies. If one needs to specify the relevant factors quickly, as in a case of food poisoning, then it would be inappropriate to form an assumption based on the traditional normative models of causal reasoning; it may be hard to go through each food item and compare food poisoning occurrence rates in the cases where the food has been eaten and in the cases not eaten (as in the case of ΔP introduced later).

The DFH is defined as the uppermost limit of the fourfold correlation coefficient phi [φ] (one of the most popular indices of linear relationship) when the value of the d-cell is infinitely large. The d-cell with an infinite divergence has the following four implications.

(1) Rarity of causes and effects
(2) Cognitive economy
(3) Monopolar causality
(4) Asymmetry between occurrence and non-occurrence

TABLE 8.2 A 2 × 2 contingency table representing covariation information between a candidate cause and a target effect

	E	$\neg E$
C	a	b
$\neg C$	c	d

Note: C and $\neg C$ represent the occurrence and non-occurrence of a candidate cause, and E and $\neg E$ represent the occurrence and non-occurrence of a target effect, respectively.

The implication that we wish to emphasise most is the fourth, which will be discussed starting in the section *Asymmetry: focalisation and negation* to clarify the distinction between A- and B-frames. For now, we will discuss the first three. Rarity of causes and effects forms the premise for the DFH. The rarer the C and E are (i.e. the lower $P(C)$ and $P(E)$ are), the greater the relative size of d will be. Therefore, insofar as a rarity assumption is established, an assumption that d is infinite serves as an effective approximation. In other words, disregarding the d-cell thus provides, as it were, a 'corner-cutting' measure.

Cognitive economy is about the trade-off between cognitive resources and calculation accuracy. Hattori and Oaksford (2007) used a computer simulation to investigate the adaptive aspect of the d-cell disregard. Assuming rarity, it is possible to calculate, disregarding the d-cell, the correlation of two events to a certain degree of accuracy. Indeed, d-cell disregard has the advantage of easing the memory load, as there is no need to retain in memory a vast number of d-cell cases. Additionally, since there is no extra memory burden, it will be possible to divert more resources to samples of other cells, which will enable a more precise estimation. In short, rather than being simply a corner-cutting measure, the d-cell disregard may be an effective heuristic device.

Monopolar causality refers to single polarity of the scale of causation (from zero to complete effectiveness). In the relevance mode of thinking (with the A-frame), the goal is to distinguish relevant events from irrelevant ones rapidly. From this point of view, absence of relevance is the default (i.e. the default is that the two events are independent of each other), and the concern is how much covariation can be observed with this default. As will be discussed later, the A-frame, by shining the spotlight on the occurrence of an event, concerns itself with the efficient detection of correlation. Therefore, it ignores negative causality heuristically. To put it another way, the A-frame is employed when there is no urgent need to detect a negative effect. In the food poisoning example, there is an urgent need to identify the item that caused the food poisoning, but there is not much need to identify an item that raises resistance to food poisoning. In accord with the A-frame, the DFH output is monopolar since the sign of phi (φ) is determined by $(ad - bc)$ and is positive when the d-cell diverges to infinity. In other words, its concern is on whether a relationship exists between C and E and, if so, how strong this relationship is. The implication of this is that an index based on such thinking does not directly deal with preventive causes. The index does not detect whether the existence of C prevents the occurrence of E.

Some authors (e.g. Lu et al., 2008) have criticized DFH as non-normative and unable to detect preventative causes. But this criticism is beside the point. DFH is a model of the relevance mode of thinking, that is the A-frame, which is not meant to be normative, but is rather generally reliable under changing circumstances, as we discuss later.

B-frame: intervention and commitment

ΔP is probably the simplest model for expressing the B-frame. It is defined as follows.

$$\Delta P = P(E|C) - P(E|\neg C) = \frac{a}{a+b} - \frac{c}{c+d}.$$

The idea underlining this model is as follows. The occurrence rate of E when C has occurred is compared to that when C has not occurred, and if the former is higher than the latter, then the difference indicates the degree of C's capacity to control E (Jenkins & Ward, 1965). This is a model of the differentiation mode in the sense that it differentiates the occurrence rate of E when C has occurred from that when C has not occurred.[1] Since the B-frame compares the presence and absence of C, it includes in its scope the detection of situations in which the presence of C actually inhibits the occurrence of E. Thus, in contrast to the A-frame, the B-frame provides a scale of *bipolar causality* (from completely preventative to completely generative).

The aim of employing the B-frame is to control the target effect. To enable accurate control of the target effect, it is essential to understand the causal structure. An effective method for understanding the causal structure between factors is intervention in the system – one makes C occur and observes whether E occurs as a result (e.g. Spirtes, Glymour & Scheines, 2001; Steyvers, Tenenbaum, Wagenmakers & Blum, 2003). Therefore, the B-frame has a high affinity with intervention. ΔP is not necessarily premised upon intervention, but it is the easiest-to-understand index for measuring the effect of intervening on C; it expresses the difference in terms of occurrence possibility between the cases when C is instigated (intervention) and when it is not (non-intervention). However, the B-frame's process load is greater than that of the A-frame. Therefore, initiation of a B-frame process may require some degree of commitment. In other words, a precondition of the B-frame is that the person be sufficiently motivated to invest the requisite cognitive resources. This is a testable hypothesis.

Asymmetry: focalisation and negation

The reason for the A-frame disregarding the d-cell is not simply proximity premised on rarity. It is deeply related to the assumption that causal reasoning is event-driven. When one is interested in the presence or absence of causation, one will usually focus on cases in which C and/or E have occurred. If we return to the ignition key example, the phenomenon of 'not turning the ignition key and not starting the car' does correspond to the d-cell, but a situation in which 'nothing occurs' does not trigger any supposition as to the relationship between the non-cause and the non-effect. Causation for people implies that one event has caused another, where an 'event' is thought of as an actual occurrence. Thus the concept of causation has an implicit presupposition: the default is a situation in which nothing is occurring. The 'event' that makes a contrast against this background of non-occurrence is deemed to be the cause. Of course, in the strictest sense, there is no such thing, objectively, as a situation in which nothing is occurring. But subjectively, non-occurrence is the default cognitive state, and there is a psychological asymmetry between occurrence and non-occurrence.

Logically, occurrence and non-occurrence have a complementary relationship. If we represent X as a hypothetical proposition, then $\neg X$ is its logical negation.

Sometimes we can find an affirmation by substituting a Y for $\neg X$, but this is not always possible, and the relationship between an affirmation and its negation is not psychologically symmetrical. Ordinary people do not usually pay attention to what does not occur. An 'exception that proves the rule' is Sherlock Holmes, who was brilliantly not like an ordinary person in his observation about 'the dog that didn't bark', which helped him to solve a mystery that baffled everyone else. Gilbert (2006) identified a lack of attention toward non-occurring things as a cause of our *illusions of foresight*, and he discussed an interesting case concerning pigeons, as well as the results of studies by Tversky (1977) and Shafir (1993), which we will come to later. Jenkins and Sainsbury (1969, 1970) demonstrated that, in the training of pigeons, there is asymmetry between the presence and absence of a distinctive feature. Pigeons were presented with two types of key – one with a distinctive feature (e.g. a key with a dot drawn on it) and one without a distinctive feature (e.g. a key with nothing drawn on it). The study showed that discrimination learning was easier in the case where the pigeons pecked the distinctive key and received a food reward than in the case where they pecked the non-distinctive key and received a food reward. Newman, Wolff, and Hearst (1980) reported that this *feature-positive effect* can also be observed amongst university students.

The asymmetrical relationship between occurrence and non-occurrence can be equated to the perceptual phenomenon of 'figure and ground' described by Rubin (1915/1958, 1921). Generally, it is occurrence (action or affirmation) that captures the attention, while non-occurrence (non-action or negation) forms the background of occurrence. This unnoticed background phenomenon is foggy and seldom cognitively processed in any great detail. The figure and ground concept is represented symbolically in the famous Rubin's vase (Figure 8.1). Rubin (1915/1958) argued that the figure and the ground function differently from each other in the perception of shape, stating that, 'in a certain sense, the ground has no shape' (p. 194). This argument corresponds with the premise for causal reasoning, in which non-occurrence is assumed as the *cognitive default*.

A closely related point is the unfeasibility of counting the *d*-cell frequency. 'Turning the key' (C) and 'starting the car' (E) are actions, but 'not turning the key' ($\neg C$) and 'not starting the car' ($\neg E$) are states. A set of actions (*a*-cell) and a set of actions and states combined (*b*- and *c*-cells) are countable, but a set of states (*d*-cell) cannot be counted as in 'first state, second state and so on'. All one can really do is fix the time of observation at a certain time of a certain day or record the states that exist at certain fixed intervals along a timeline. This situation indicates that the *d*-cell is peculiar among the data, which forms the rational basis for assuming the *d*-cell to be the 'ground'.

So far, we have discussed the A-frame as a particular cognitive framework for causal induction, but it may be more appropriately regarded as a general cognitive framework for generating the asymmetry between affirmative and negative occurrences. This proposal is based on the fact that a similar structure can be found in a wide variety of areas, including reasoning, hypothesis testing, similarity, probability

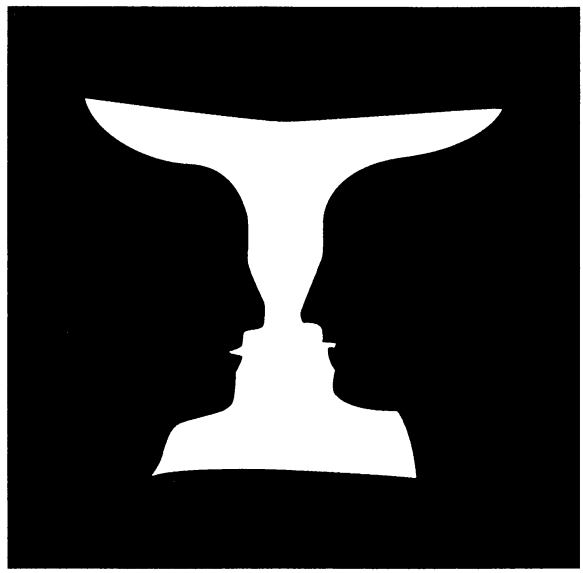

FIGURE 8.1 Rubin's vase (Rubin, 1921)

judgement, preferences, theory of mind and social inference. It is unclear at present whether these areas all use the same perceptual framework, but seeking out the commonality between them will be an important first step toward constructing an integrated theoretical model. This being the case, from this point on, we will look at how various errors and biases can be reinterpreted as instances of this kind of asymmetry.

Asymmetry in deduction

The first person to examine experimentally the asymmetry between affirmation and negation was probably Wason (1959). In his pioneering experiment, in which he used as indices the reaction time and number of errors in processing sentences, Wason demonstrated that affirmative information is easier to process than negative information. It seems simple enough that, if a sentence is true, its negation is false. For example, if 'the star is yellow' is a true affirmative sentence, then its negation, 'the star is NOT yellow,' is false. Conversely, if a sentence is false, its negation is true. There are therefore $2 \times 2 = 4$ statements: true affirmative, false affirmative, true negative and false negative. Wason investigated differences in the relative ease of processing the information of such statements. He presented participants with an illustration showing the true 'situation'. The illustration showed a number of green, red, yellow and black stars located in various positions numbered from 1 to 4. Participants were asked to select the wording

to make the sentence accurately represent the situation. The participants were provided with affirmative and negative statements. The following is an example of an affirmative statement.

$$\text{There is both } \left\{\begin{matrix}\text{Yellow}\\ \text{Green}\end{matrix}\right\} \text{ in 4 AND } \left\{\begin{matrix}\text{Red}\\ \text{Black}\end{matrix}\right\} \text{ in 3.}$$

The following is an example of a negative statement.

$$\text{There is NOT both } \left\{\begin{matrix}\text{Yellow}\\ \text{Green}\end{matrix}\right\} \text{ in 4 AND } \left\{\begin{matrix}\text{Red}\\ \text{Black}\end{matrix}\right\} \text{ in 3.}$$

In the experiment, the mean reaction time was quickest with true affirmative statements, followed in order by false affirmatives, true negatives and false negative statements. The results thus revealed that affirmative information requires less cognitive load than negative information, and in our view, this experiment clarified the asymmetry between affirmation and negation.

In the year following this experiment, Wason (1960) reported confirmation bias in his 2–4–6 task, which accentuates the asymmetry between affirmative and negative information. People tend to seek confirmation of their own hypothesis by finding instances that support it and to avoid falsification of it through the use of contradicting instances. Our attention is captured by a hypothesis supported by positive instances, and affirmative feedback about these instances. The falsity of a hypothesis and negative instances are consigned to the background. The affirmative aspects of hypotheses and their instances are the 'figure', and the negative aspects of hypotheses and instances are the 'ground'. Of particular interest is the fact that positive feedback for a positive test (Klayman & Ha, 1987) has different psychological implications than negative feedback for a negative test. Both are the cases of confirmation, but the former is 'figure' confirmation and, as such, captures attention and becomes the target of bias, while the latter is the 'ground' confirmation and, as such, does not capture attention.

The asymmetry of confirmation is laid bare by Hempel's (1945) famous ravens paradox. Here the hypothesis of interest is that 'all ravens are black' (H_1), which will be confirmed by the existence of a black raven. Similarly, the hypothesis that 'everything that is not black is not a raven' (H_2) is confirmed by the existence of an object that is not black and is not a raven, for example a white shoe or red umbrella. But these two hypotheses are logically equivalent to each other, Hempel argued, by contraposition in classical logic, and he concluded that hypothesis H_1 will be equally confirmed by a white shoe and a red umbrella. The following two thinking approaches resolve this apparent paradox, and they both can be related to the concept of 'figure and ground'.

The first approach resolves the paradox through a Bayesian model. In this model, hypothesis testing is regarded as a comparison of likelihood between competitive models (McKenzie & Mikkelsen, 2000; Oaksford & Chater, 2007). If a black raven (D) is observed, this could mean two things. It could mean that hypothesis H_1 is true and that the observation of D is supporting evidence of the hypothesis. At the same time, it could also mean that D is observed by chance in spite of the falsity of H_1 (e.g. raven and blackness are independent of each other [H_0]). If the probabilities of ravens and of black things are both small, then the former (i.e. H_1 is true) is much more likely to be the case because it is improbable that D is observed by chance in such a case. In this context, the probability of ravens (or black things) refers to the probability that, if you randomly pick out a thing, it will be a raven (or a black thing). This world contains a vast diversity of things, so the probabilities of ravens and black things are very small; in other words, ravens and black things have *rarity*. For hypothesis H_1, the expected information gain of D will be much higher than that of, say, an observed white shoe. In other words, the observation of D will raise the certainty factor of the hypothesis, whereas the observation of a white shoe would hardly raise it at all. Discovering a black raven would have considerable significance, but discovering a white shoe would have negligible significance. 'Things that have rarity' (ravens and black things) are the 'figure', while everything that is not raven or black is the 'ground'.

The second approach is based on recent psychological studies of the indicative conditional of natural language. This research has supported the hypothesis that this conditional, *if p, then q*, is a *probability conditional*, or *conditional event*, and not the material conditional of classical logic (Baratgin, Over & Politzer, 2013, 2014; Baratgin & Politzer, 2016; Evans & Over, 2004; Over & Baratgin; 2016, this volume; Politzer & Baratgin, 2016). The probability of the probability conditional, P(*if p, then q*), is the conditional probability $P(q \mid p)$, which is not the same as $P(\textit{not-p} \mid \textit{not-q})$, and so contraposition fails for the probability conditional: *if p, then q* is not logically equivalent to *if not-q, then not-p*. For a particular object, the probability of 'if raven (R), then black (B)' is equal to $P(B \mid R)$, the probability that the object is black given that it is a raven. This can be determined by finding the proportion of black ravens out of all ravens, which can be very different from the proportion of non-ravens among non-black objects, $P(\neg R \mid \neg B)$ (Fitelson & Hawthorne, 2009). White shoes and red umbrellas are totally irrelevant here; we must investigate the ravens themselves. As long as we do not investigate ravens, either by searching out every single one or by using a sample, we will not be able to obtain the data on the relevant probability. Thus, the precondition (raven) and the post-condition (black thing) are the 'figure', and everything else is the 'ground'.

Asymmetry in attributes

Until now, we have discussed the logical affirmation and negation of events or things themselves, but similar arguments can be made with regard to the characteristics associated with things (i.e. attributes). A target event or thing has innumerable

attributes, but an attribute that captures attention is given weight in processing. As a result, the same target object can have different cognitive processing in different contexts. For example, Levin (1987) reported that ground beef labelled '75% lean' was rated more positively on scales for 'high quality–low quality' and 'greaseless–greasy' compared to ground beef labelled '25% fat'. With lean as the 'figure' and fat as the 'ground', the affirmative attribution of lean is emphasised. Conversely, if fat is viewed as the 'figure', it will be the negative attribution of fat that is emphasised. In a similar vein, Sanford, Fay, Stewart and Moxey (2002) reported that yoghurt bearing the indication '25% fat' will be judged as less healthy than yoghurt bearing the indication '75% fat free'.

Judgements about similarity and preference are also influenced to a great extent by focalisation, viewpoint, situation and context. The concept of similarity was critically examined by the philosopher Goodman (1972). Goodman cited eight viewpoints, the most important of which concerns focalisation. Goodman illustrated his point using an allegory of baggage at an airport checking station (p. 445). Onlookers may focus on shape, size, colour and material; the pilot will be concerned with weight and the passengers with destination and ownership. Consequently, which baggage appears similar depends on the context of the judgement and on who is making the comparison. In short, 'circumstances alter similarities.' Therefore, Goodman argues, inasmuch as the concept of similarity does not provide a place for context, it has no use as a tool for philosophical analysis.

Tversky (1977) discussed the same issue, and it was he who constructed one of the most influential models of similarity. In Tversky's model, similarity judgements are based on features, and by this model, similarities and differences do not complement one another. Usually, the more similar *A* and *B* are, the greater their ratings of 'similarity' are. One would expect that their ratings of 'difference' would, conversely, be lower and that adding the two together would result in a fixed figure (such as 100), but Tversky demonstrated that this is not always the case. If the objects being compared are well known, they will have many common and distinctive features. If the objects are not so well known, then they will have few common and distinctive features. It is common features that capture the attention in similarity judgements, and distinctive features that capture attention in difference judgements. Therefore, one can expect that, if the items compared are a *prominent pair* (a pair of items that are well known), they are more likely to be judged as similar to each other – and also different from each other – than if they were a *non-prominent pair*. In Tversky's (1977) study, half of the participants (i.e. a *similarity* group) were presented with two pairs of countries (a total of four countries) and asked to judge which of the two pairs were more *similar* to each other. The rest of the participants (i.e. a *difference* group) were given the same stimulus material but asked to judge which pair were more *different* from each other. In both the similarity group and the difference group, around 70% of the participants from Israeli colleges selected 'West Germany – East Germany' as opposed to 'Ceylon – Nepal'. Thus, 'West Germany – East Germany' were judged as similar, and they were also judged as different. The same results have been obtained for other prominent pairs.

Similarity judgements are not the only factor influencing the way an object's attributes are processed. Shafir (1993) observed the same phenomenon in decision making. He found that an *enriched* option (an option with more positive and negative features) is more likely to be targeted for both selection and rejection than an *impoverished* option (an option with moderate features). Participants were asked to imagine that they were serving on the jury of an only-child sole-custody case. They were presented with the profiles of two parents and were asked to answer the questions, 'To which parent would you award sole custody of the child?' and 'Which parent would you deny sole custody of the child?' The results showed that the parent representing the enriched option was more likely to be targeted in both questions. When selecting which option to 'choose', attention is drawn to positive features, which form a basis for choosing; when rejecting something, attention is drawn to negative features, which form a basis for rejecting.

Asymmetry in meta-representation

The asymmetry between affirmation and negation is also recognised in higher-order inference (i.e. belief about belief). Birch and Bloom (2003) demonstrated that asymmetry bias traces back to the preschool psychological state in a modified version of the Smarties task (Perner, Leekam & Wimmer, 1987). They presented 3-, 4- and 5-year-old children with two sets of toys, each with an object inside. The children were told that one of the toys was familiar to a puppet introduced to the children as the experimenter's friend, and that the other was unfamiliar to the puppet. The children were shown the toy contents half of the time and not shown the toy contents the other half (i.e. puppet's familiarity/unfamiliarity with toy × child's knowledge/ignorance of contents). The children were asked to judge whether the puppet would know what was in the toys, to test their ability to infer the belief of another 'mind' accurately. The 3- and 4-year-old children, when they knew the contents, tended to assume that the puppet would also know them, regardless of whether the puppet was familiar with the toy. When they did not know the contents, they tended to assume correctly that the puppet would know if it was familiar with the toy, and these children were unlikely to assume incorrectly that the puppet would know if it was not familiar with the toy. Thus, preschool children assume that the things they know are also known by others, but on the other hand, they do not assume that the things they do *not* know are also *not* known by others.

These results are said to be evidence of what is called the *curse of knowledge*. The curse of knowledge was originally proposed in the field of economics and can be described as follows: 'Better-informed agents are unable to ignore private information even when it is in their interest to do so; more information is not always better' (Camerer, Loewenstein & Weber, 1989, p. 1232). The inability of children aged 3 or under to solve false belief tasks is indicative of their tendency to assume that the things they know are known by others even if there is no way that others could know them. All the way back to Piaget, such a tendency has traditionally been understood as *egocentrism*, but it is gradually becoming clear that children are

equally poor at recalling their own past psychological states (Gopnik & Astington, 1988), and that adults also have difficulties at inferring the psychological states of others and their own past psychological states (Birch & Bloom, 2007). Based on these findings, Birch and Bloom (2003) provided the curse of knowledge explanation of this phenomenon.

The curse of knowledge is a better explanation than egocentrism in two respects. First, it can account for the asymmetry in the study's results. The study found that the children 'were biased by their knowledge when attempting to appreciate the perspective of someone more ignorant than themselves, but were not biased by their ignorance when attempting to appreciate the perspective of someone more knowledgeable than themselves' (Birch & Bloom, 2003, p. 285). Egocentrism alone cannot account for this finding. Second, the difficulty of inferring someone else's beliefs, a 'theory of mind' problem (Premack & Woodruff, 1978), can be understood comprehensively on the same level as various other (adult) cognitive biases, including social cognition. Birch and Bloom cited a number of biases related to the curse of knowledge, including the *hindsight bias* (Fischhoff, 1975), the *spotlight effect* (Gilovich, Medvec & Savitsky, 2000), the *illusion of transparency* (Gilovich, Savitsky & Medvec, 1998), and the *false consensus effect* (Ross, Greene & House, 1977).

It is possible that at least some of these belief-related biases could be understood integrally through the same cognitive framework – namely, higher-order frames. Our attention is drawn to the things that we know and believe, whereas we tend not to focus on things that we do not know or believe. Beliefs (including false beliefs) are often the 'figure', but non-beliefs cannot easily become the 'figure'. A belief can be represented as a second-order predicate, for example, 'John believes X is at Y': Believe (*John*, Exist(*X*, *Y*)). There are also nested beliefs, also described as second-order beliefs, such as 'Mary believes John believes it is here.' Just as there are higher-order beliefs, it is conceivable that there are also higher-order frames, and that it is from these higher-order frames that the asymmetry between affirmation and negation is generated. There will be much value in empirically verifying whether the phenomenon occurs in other biases Birch and Bloom cited, such as hindsight bias. This bias refers to the tendency to look back upon an event that has occurred and assume that the event was a natural, expected result. When something occurs (e.g. someone dies), people often feel like they predicted the event (i.e. death) would occur all along. However, when nothing occurs, they do not get the equivalent feeling – they do not feel that they predicted all along that nothing would occur. Thus, asymmetry exists between positive and negative beliefs. It can be argued, therefore, that the curse of knowledge is a product of asymmetric higher-order frames. This is a hypothesis that is sufficiently testable.

Trade-offs and frame-switching

We have seen how the A-frame is related to various biases. If these biases are to be considered robust, it follows that they must have some kind of epistemological

utility in the A-frame. The most likely candidate is cognitive load reduction. We already explained cognitive load reduction in relation to the DFH case. Another important point is that cognitive load may be related to the structure by which the A-frame resolves the *frame problem* (McCarthy & Hayes, 1969; Pylyshyn, 1987).

The frame problem is fundamentally important and as yet unresolved in artificial intelligence (AI) studies. It concerns the fundamental difficulty of an agent with finite processing capacity with problems in a complex environment. The initial worry was that, if the agent in such an environment carries out some action or if something external to the agent occurs, a vast number of circumstances remain unaffected, but to give an account of every one of these circumstances would entail an explosion in computations and the amount of data. So far AIs are not as capable of disregarding irrelevant (or less important) information as human beings, for the AIs work exclusively in a B-frame mode. In this sense, studying the dual frames in higher cognition could contribute to intelligent machine developments.

The A-frame, of course, is not omnipotent. It is heuristic and, under certain conditions, will fail to perform. Sometimes 'the dog that didn't bark' is significant! The various cognitive biases we have explored in this chapter are examples of such failures. Given that the A-frame is one perspective, someone who wants to reduce biases may pose a question: 'Can the biases be reduced by changing the perspective?' A suggestion for answering this question can be found in the results of the experiment introduced in the *Asymmetry in deduction* section on the 2–4–6 task. Tweney and colleagues (1980, Experiment 4) discovered that modifying the form of the questions dramatically improved correct answer rates. In a modified version of the 2–4–6 task called the DAX-MED version, triples are assigned to one of two complementary categories: DAX ('three ascending numbers') or MED (other combination). In this version, the feedback the participants received about each triple was not whether the triple conforms to the rule (affirmation vs. negation) but rather one of two alternatives: whether it is categorised as DAX or MED. The correct answer rate in this version was four times higher than in the initial study (60% vs. 15%). The same finding has been obtained in a number of other studies (e.g. Tukey, 1986; Wetherick, 1962; Wharton, Cheng & Wickens, 1993). This task is a dual-goal task, and goal complementarity has been suggested as the key reason for the results (Gale & Ball, 2006). The implication of the results is that it is possible for A-frame use to be diminished by some factors. The frame one employs when categorising objects (e.g. triples) depends on whether one sees them as either *X* (e.g. ascending numbers) or *not-X* (e.g. others) or complementary alternatives (e.g. DAX vs. MED). This hypothesis is also testable.

Conclusion

In this chapter, we have proposed a theoretical foundation for a distinction between what we have called A/B-frames. Beginning with the relatively low-level cognitive function of visual perception, we went on to an examination of higher-order cognitive functions – like deduction, induction, judgement, decision making and

problem solving – and then of particular phenomena covered by social cognition and cognitive development. Whether all of these phenomena are truly founded on a common cognitive mechanism must await the findings of future research. The significance of the framework proposed here must also be judged by the impact of its explanations and the extent to which it expands research, generating testable hypotheses. It is necessary to verify the claim that this diversity of phenomena shares commonality. We must see whether it is possible to predict that what holds for one phenomenon will also be true for as-yet-undiscovered aspects of another phenomenon. For example, Birch and Bloom (2004) proposed inhibitory control as the neural basis for the asymmetry in the curse of knowledge, but we should go beyond false belief tasks and consider visual tasks and the illusion of transparency, to determine whether they have the same neural basis. But most of all, it is necessary to verify empirically the psychological reality of A/B-frames.

Acknowledgment

This study was supported by JSPS-ANR CHORUS Program J121000148 and JSPS KAKENHI Grant 15H02717.

Note

1 The normative model most often proposed in research to date is the differentiation mode model. For example, there is a model that takes into consideration the role of alternative cause with regard to ΔP (Cheng, 1997; Lu et al., 2008); a model that incorporates weighting parameters into ΔP (Anderson & Sheu, 1995; Rescorla & Wagner, 1972; Wasserman, Elek, Chatlosh, & Baker, 1993); and a causal support model, which differentiates between the presence and absence of a causal relationship between the candidate cause and the target effect (Griffith & Tenenbaum, 2005). The present study does not attempt to compare these models. Instead, it focuses on the ΔP model as it is the simplest and most well known in our view.

References

Anderson, J. R., & Sheu, C.-F. (1995). Causal inferences as perceptual judgements. *Memory & Cognition, 23*, 510–524.

Baratgin, J., Over, D. E., & Politzer, G. (2013). Uncertainty and de Finetti tables. *Thinking & Reasoning, 19*, 308–328.

Baratgin, J., Over, D. E., & Politzer, G. (2014). New psychological paradigm for conditionals and general de Finetti tables. *Mind and Language, 29*, 73–84.

Baratgin, J., & Politzer, G. (2016). Logic, probability and inference: A methodology for a new paradigm. In L. Macchi, M. Bagassi & R. Viale (Eds.), *Cognitive unconscious and human rationality* (pp. 119–142). Cambridge, MA: MIT Press.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science, 14*, 283–286.

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences, 8*, 255–260.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false beliefs. *Psychological Science, 18*, 382–386.

Camerer, C. F., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97*, 1232–1254.

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review, 104*, 367–405.

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

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fitelson, B., & Hawthorne, J. (2009). How Bayesian confirmation theory handles the paradox of the ravens. In E. Eells & J. Fetzer (Eds.), *The place of probability in science* (pp. 247–275). Netherlands: Springer.

Gale, M., & Ball, L. J. (2006). Dual-goal facilitation in Wason's 2–4–6 task: What mediates successful rule discovery? *The Quarterly Journal of Experimental Psychology, 59*, 873–885.

Gilbert, D. T. (2006). *Stumbling on happiness*. London: Harper Perennial.

Gilovich, T., Medvec, V. H., & Savitsky, K. (2000). The spotlight effect in social judgement: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology, 78*, 211–222.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: Biased assessments of other's ability to read one's emotional states. *Journal of Personality and Social Psychology, 75*, 332–346.

Goodman, N. (1972). *Problems and projects*. Indianapolis, IN: Bobbs-Merrill.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26–37.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology, 51*, 334–384.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review, 116*, 661–716.

Hattori, M., & Oaksford, M. (2007). Adaptive non-interventional heuristics for covariation detection in causal induction: Model comparison and rational analysis. *Cognitive Science, 31*, 765–814.

Hempel, C. G. (1945). Studies in the logic of confirmation (I.). *Mind, 54*, 1–26.

Jenkins, H. M., & Sainsbury, R. S. (1969). The development of stimulus control through differential reinforcement. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 123–161). Halifax, Nova Scotia, Canada: Dalhousie University Press.

Jenkins, H. M., & Sainsbury, R. S. (1970). Discrimination learning with the distinctive feature on positive or negative trials. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 239–273). New York: Appleton-Century-Crofts.

Jenkins, H. M., & Ward, W. C. (1965). Judgement of contingency between responses and outcomes. *Psychological Monographs: General and Applied, 79*, 1–17. (Whole No. 594).

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Levin, I. P. (1987). Associative effects of information framing. *Bulletin of the Psychonomic Society, 25*, 85–86.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955–984.

Manktelow, K. I. (2012). *Thinking and reasoning*. Hove, UK: Psychology Press.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of the artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence* (Vol. 4, pp. 463–502). Edinburgh, UK: Edinburgh University Press.

McKenzie, C. R. M., & Mikkelsen, L. A. (2000). The psychological side of Hempel's paradox of confirmation. *Psychonomic Bulletin & Review, 7*, 360–366.

Newman, J. P., Wolff, W. T., & Hearst, E. (1980). The feature-positive effect in adult human subjects. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 630–650.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608–631.

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

Over, D. E., & Baratgin, J. (2016). The 'defective' truth table: Its past, present, and future. In E. Lucas, N. Galbraith & D. E. Over. (Eds.), *The thinking mind: The use of thinking in everyday life*. Hove, UK: Psychology Press.

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology, 5*, 125–137.

Politzer, G., & Baratgin, J. (2016). Deductive schemas with uncertain premises using qualitative probability expressions. *Thinking & Reasoning, 22*, 78–98.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioural and Brain Sciences, 1*, 515–526.

Pylyshyn, Z. W. (1987). *Robot's dilemma: The frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Ross, L., Greene, D., & House, P. (1977). The 'false consensus effect': An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.

Rubin, E. (1915/1958). Figure and ground. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 194–203). Princeton, NJ: D. Van Nostrand. English translation of key sections from Rubin's dissertation.

Rubin, E. (1921). *Visuell wahrgenommene Figuren: Studien in psychologischer analyse [Visually perceived figures: Studies in psychological analysis]*. Kobenhavn: Gyldendalske Boghandel.

Sanford, A. J., Fay, N., Stewart, A., & Moxey, L. (2002). Perspective in statements of quantity, with implications for consumer psychology. *Psychological Science, 13*, 130–134.

Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition, 21*, 546–556.

Shanks, D. R., Lopez, F. J., Darby, R. J., & Dickinson, A. (1996). Distinguishing associative and probabilistic contrast theories of human contingency judgment. In D. R. Shanks, K. Holyoak & D. L. Medin (Eds.), *Causal learning* (pp. 265–311). San Diego, CA: Academic Press.

Spirtes, P., Glymour, C., & Scheines, R. (2001). *Causation, prediction, and search* (Second edition). New York: Springer.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science, 27*, 453–489.

Tukey, D. D. (1986). A philosophical and empirical analysis of subjects' modes of inquiry in Wason's 2–4–6 task. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 38*, 5–33.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327–352.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt, C. R., Gross, K. A., et al. (1980). Strategies of rule discovery in an inference task. *The Quarterly Journal of Experimental Psychology, 32*, 109–123.

Wason, P. C. (1959). The processing of positive and negative information. *The Quarterly Journal of Experimental Psychology, 11*, 92–107.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology, 12*, 129–140.

Wasserman, E. A., Elek, S. M., Chatlosh, D. L., & Baker, A. G. (1993). Rating causal relations: Role of probability in judgements of response-outcome contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 174–188.

Wasserman, E. A., Kao, S.-F., Van Hamme, L. J., Katagiri, M., & Young, M. E. (1996). Causation and association. In D. R. Shanks, K. Holyoak, & D. L. Medin (Eds.), *Causal learning* (pp. 207–264). San Diego, CA: Academic Press.

Wetherick, N. E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology, 14*, 246–249.

Wharton, C. M., Cheng, P. W., & Wickens, T. D. (1993). Hypothesis-testing strategies: Why two goals are better than one. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 46A*, 743–758.

White, P. A. (2003). Making causal judgments from the proportion of confirming instances: The pCI rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 710–727.