PATTERNS, NOISE, AND BELIEFS

LAJOS L. BRONS

Nihon University & Lakeland University, JAPAN
mail@lajosbrons.net

Abstract. In "Real Patterns" Daniel Dennett developed an argument about the reality of beliefs on the basis of an analogy with patterns and noise. Here I develop Dennett's analogy into an argument for descriptivism, the view that belief reports do no specify belief contents but merely *describe* what someone believes, and show that this view is also supported by empirical evidence. No description can do justice to the richness and specificity or "noisiness" of what someone believes, and the same belief can be described by different sentences or propositions (which is illustrated by Dennett's analogy, some Gettier cases, and Frege's puzzle), but in some contexts some of these competing descriptions are misleading or even false. Faithful (or truthful) description must be guided by a principle (or principles) related to the principle of charity: belief descriptions should not attribute irrationality to the believer or have other kinds of "deviant" implications.

Keywords: Beliefs ● mental content ● propositional attitude reports ● descriptivism ● Frege's puzzle ● principle of charity.

1. Belief reports

The traditional view on beliefs and belief reports is that a belief report specifies the content of a belief and that that content is (something very much like) a proposition. In "On Sense and Reference" (1892), Frege showed that — together with some other widely shared assumptions about language — this traditional view has paradoxical implications, however. Frege's "puzzle" is that, if x believes that y is F under one name (for y) but not under another (because x doesn't realize that the two names refer to the same individual), then x holds contradictory beliefs (or at least appears to do so). Frege's own example was about Hesperus and Phosphorus, but modern illustrations are typically about Superman: Lois Lane believes that Superman is strong and that Clark Kent is not strong, but (unbeknownst to Lois Lane) Clark Kent is Superman, and thus, Lois Lane believes that the same individual is both strong and not strong.

   Frege's solution to this puzzle was to give up direct reference and semantic innocence, the aforementioned "other widely shared assumptions about language". (Direct reference is the idea that singular terms point at things in the world directly. Se-

mantic innocence is the assumption that the meaning of words and phrases remains constant across linguistic contexts.) According to Frege, terms have both a sense and a reference. While "Superman" and "Clark Kent" *refer* to the same individual, they do not have the same sense, and therefore, the terms "Superman" and "Clark Kent" in Lois Lane's two beliefs are not semantically innocent,[1] and reference is mediated by sense (and thus not direct).

A problem very similar to Frege's puzzle was discussed by Quine (1976 [1956]) in the context of quantification into beliefs. Quine's "Ortcutt" case is formally identical to Frege's puzzle except that the two beliefs involve references to the same individual by means of different descriptions rather than different names. "Ralph believes that the man in the brown hat is a spy" and "Ralph does not believe that the man seen at the beach is a spy" (p.187), but unbeknownst to Ralph, both descriptions refer to the same man named "Ortcutt".

David Kaplan (1968) proposed a solution to Quine's quantification problem that was strongly influenced by Frege and that also solves Frege's puzzle. A belief of x that y is F if represented as α is formalized as "$\mathbf{R}(\alpha, y, x) \;\wedge\; x\mathbf{B}^\ulcorner\alpha$ is $F^\urcorner$" in which the two-place predicate $\mathbf{B}$ stands for "believes that" and the three-place predicate $\mathbf{R}$ for representation. Lois Lane has two representations of the same individual, one as Superman and one as Clark Kent, and has different beliefs about these two representations. Thus, it is the case that $\exists\alpha[\mathbf{R}(\alpha,s,l) \;\wedge\; l\mathbf{B}^\ulcorner\alpha$ is strong$^\urcorner]$ and $\exists\alpha[\mathbf{R}(\alpha,s,l)\wedge\neg(l\mathbf{B}^\ulcorner\alpha$ is strong$^\urcorner)]$ (in which s stands for Superman/Clark Kent and l for Lois Lane), but that does not imply a contradiction. (See Kaplan 1968, pp.206–7 for his analogous treatment of the Ortcutt case.)

Kaplan defines representation (in the here relevant sense) as follows: α represents y to x — symbolized as "$\mathbf{R}(\alpha, y, x)$" — if and only if α denotes y, α is a name of y for x, and α is sufficiently vivid.[2] What α is a name of (for x) is determined by α's "genetic character". "The genetic character of a name in a given person's usage will account for how he *acquired* the name, that is how he heard of such a thing and, if he believes that such a thing exists, how he came to believe it" (p.200). A representation, then, is something like a "mental file" or "mental dossier" x has (formed) of y, and like Frege's senses, such representations mediate between names (or descriptions, as in the Ortcutt case) and referents. Nevertheless, the two notions should not be confused: mental files or dossiers are *not* Fregean Senses.

Frege distinguished the sense of a term or name from its representation (*Vorstellung*). "The latter is subjective, private and psychological, the former is objective and communicable" (Kremer 2010, p.260). Mental files or dossiers are subjective, private and psychological and, thus, *Vorstellungen* (representations). The distinction matters for how one conceives of the content of beliefs. According to Frege, the sense of a sentence is a thought (*Gedanke*), which is "not the subjective activity of thinking, but its objective content, which is capable of being the common property of many [thinkers]"

(1892, p.32n; my translation). The senses of thoughts are constructed out of the building blocks provided by the senses of words (Kremer 2010, p.289), and both kinds of senses are equally objective and sharable. Fregean thoughts (*Gedanken*), then, are very similar to propositions: they are objective, truth-apt, sharable, structured, and stand in logical relations to each other. (See also Taschek 2010.) Hence, while Frege gives up semantic innocence and direct reference, he (more or less) adheres to the aforementioned traditional view of beliefs and belief reports. Whether the same is the case for a theory that substitutes mental files for senses as suggested by Kaplan's account is debatable, however. Propositions are typically assumed to be sharable and, therefore, must have sharable building blocks, but the genetic character of a name or term is inherently subjective and unique.

While the Fregean approach to solving the puzzle by giving up direct reference and semantic innocence might work, many philosophers find the prize too high to pay, and many other solutions (or attempts at solutions, at least) have been proposed. (For overviews, see Richard 1997, Brogaard 2007, or McKay and Nelson 2010.) A prominent kind of solution appeals to hidden indexicals such as contexts or ways of believing. For example, Mark Crimmins and John Perry (1989) argue that Lois Lane's two beliefs have the same content, but differ in *how* she believes that content. And according to Mark Richard (1990), what makes a belief report true is that the reporting sentence faithfully represents what the believer believes, and:

> What counts as faithful representation varies from context to context with our interests and expectations. Context places certain restrictions on what can represent what. Sometimes they are very specific. . . Sometimes context is silent. . . , and expressions are free to represent any expressions with which they corefer. (pp.3–4)

Whether such solutions succeed and to what extent they respect (both) direct reference and semantic innocence is controversial. It seems that every proposed solution conflicts with at least one intuitively plausible principle. Nevertheless, all of the solutions mentioned (as well as most of those not mentioned) agree in at least one respect: they hold on to the aforementioned traditional view — that is, they assume that belief reports specify the contents of beliefs, and that those contents are (very similar to) propositions. According to Kent Bach (1997) these assumptions are the root of the problem, however. A "belief report does not quite what it appears to do, namely, say what someone believes. That is, it does not specify what the person believes but merely describes it" (p.215), and therefore, the "specification assumption" is false. Bach argues that "belief reports are semantically incomplete" — that is, they are "not true or false independently of context" (p.238).

Largely independent from the dispute about belief reports and Frege's puzzle there has been simultaneous debate about the metaphysics of beliefs (although there

are obvious links between the two). In the metaphysical debate representationalists spar with dispositionalists about the nature of beliefs, and together they confront eliminativists and instrumentalists on the question of the existence of beliefs. (For an overview, see Schwitzgebel 2015.) In the latter (sub-) debate, Daniel Dennett took a position that he himself considered a kind of realism, but that is more commonly labeled "instrumentalist". The main paper outlining his position is "Real Patterns" (1991). The argument in that paper builds on the analogy that gave it its title: patterns and their reality or unreality.

While Dennett used his pattern analogy in an argument about the existence of beliefs, in this paper I will use it to develop an account of belief reports. That account agrees with Bach's descriptivism (i.e. it holds that belief reports merely describe beliefs), but fills in various details (mainly about the content of beliefs on which Bach is mostly silent) in a way that bears similarities to Kaplan's neo-Fregeanism as well as to hidden-indexicalism or contextualism. In the next section, I will unpack the key element of Dennett's analogy, "noise". After that, section 3 discusses the formal nature of belief reports, argues for a principle to decide between competing, more or less "noisy" — that is, more or less specific or detailed — belief descriptions, the *principle of (avoiding) deviant implications* (PDI), and illustrates that principle with a kind of Gettier case. The much shorter fourth section suggests a solution to Frege's puzzle based on the descriptivist view of belief reports defended in section 3, and a solution of Kripke's puzzles by means of PDI. The final section summarizes this paper's main claims, compares them with Bach's descriptivism, and concludes that PDI and the solution to Frege's puzzle in section 4 are really specific versions of the principle of charity. The remainder of this opening section introduces the pattern analogy.

Daniel Dennett's (1991) argument in "Real Patterns" starts with the introduction of a pattern named "*bar code*" that consists of a row of nine 10×10-dot squares alternately black and white. To this base pattern more or less "noise" is added to create the six "frames" shown in figure 1.[3] The amount of noise added ranges from 1% in *D* to 50% in *F*. (Noise ratios are shown in the figure.)

Figure 1: A recreation of Dennett's six "frames" (with noise ratios)

In what these frames are an analogy of it is not the case that the pattern is there first and the noise added later, however. Rather, the frames — as such, *with* the noise — are there first, and the question is whether the pattern is there in those frames; or whether, when, and/or to what extent the pattern can be said to describe (some of) those frames.

Dennett argues that "a pattern exists in some data — is real — if *there is* a description of the data that is more efficient than the bit map, whether or not anyone can concoct it" (p.34). However, we can take different attitudes towards patterns and the noise "obscuring" them at different times, depending on circumstances, purposes, and so forth.

> Sometimes we care about exact description or reproduction of detail, at whatever cost. From this perspective, a real pattern in frame A is *bar code with the following exceptions: 7, 8, 11*,... At other times we care about the noise, but not where in particular it occurs. From this perspective, a real pattern in frame A is *bar code with 25% noise*. And sometimes, we simply tolerate or ignore the noise. From this perspective, a real pattern in frame A is simply: *bar code*. (p.35; italics in original)

The key question for Dennett is whether the pattern, *bar code*, is *really* there in frame A. Beliefs are subject to "noise" in a way similar to the patterns in figure 1, and consequently, a careful consideration of this key question is intended to shed light on the question about the reality of beliefs. Dennett further develops his argument partially by means of another analogy (the "game of life") and by contrasting his views to those of Fodor, Churchland, Davidson, and Rorty. His conclusion is that beliefs (like patterns) are real, but also "up to the observer" (p.49), although this apparent relativism is constrained by pragmatic considerations similar to those in, for example, Quine's "On What There Is" (1964 [1948]) or *Word and Object* (1960, especially chapter 1, §6).

Regardless of whether *bar code* "is really there" in frame A (and of what exactly that is supposed to mean), we can also ask the question whether, when, and/or to what extent it is correct (or appropriate, or faithful) to *describe* frame A as *bar code*. That — rather than the existential question — is the key question here, in this paper. And to answer that question we need to start with an analysis of the "noise".

2. Noise

"Noise" is an ambiguous term, but much of this ambiguity can be contained by distinguishing four "stages" in the term's history. If we skip etymology (which links the word to "nausea"), the oldest and most basic sense of "noise" is inharmonious and/or unwanted (and often loud) sound. Street noise and factory noise are examples of

noise in this sense. After the invention of recording technology, the term gained a new use in reference to hiss and other accidental and unwanted artifacts in recordings of sound. Let's call this kind of noise "audio noise". From the 1940s a new technical notion of noise to describe various hissing sounds developed. Noise in this sense is a random signal in (usually) all (audible) frequencies, although the intensity per frequency depends on the kind — or color — of noise. White noise is the best known example. The fourth stage originates in the use of "noise" as metaphor. Noise in this sense is any disturbance or random fluctuation in a sign or signal. Let's call this kind of noise "signal noise".

The notion of signal noise is rooted in an apparent analogy to audio noise, but differs from it in a fundamental way. Signal noise is a disruption of an originally pure and undisrupted signal. The signal was there first, and the noise only came later. In the case of audio noise, however, there never was a "pure" and undisrupted signal (except, perhaps, in some electronic music). Audio noise is an inherent and unavoidable part of any sound recording — it is there from the start.

The noise in Dennett's pattern analogy can be understood as a kind of signal noise or as analogous to audio noise. If it is understood as signal noise, then the pattern *bar code* is there first, and the noise is a later disruption of that pattern. (This is, of course, how the frames in figure 1 are created, but that does not necessarily matter for how the analogy is supposed to work.) If it is understood as analogous to audio noise, then the pattern *bar code* is an idealization that never existed separate from the noise. Dennett's argument depends on the second understanding of noise in the analogy. If the pattern *bar code* is already there (and merely disrupted), then it doesn't make much sense to ask — as Dennett does — whether the pattern is "really there". The point of the analogy (or part of the point, at least), is that there is no "signal" separate from the noise. The frames are what is given, and the pattern is merely *seen in* the frames. What we see as the signal is an idealization, and Dennett's question is about the reality of that idealization. What concerns us here, however, is not the idealization's reality, but its truthfulness, faithfulness, or accuracy.

2.1. "Noisy" beliefs

The frames in figure 1 are analogous to beliefs. If Hanako believes that John drives a Toyota, then the proposition "John drives a Toyota" is analogous to the pattern *bar code*. "John drives a Toyota" *describes* Hanako's belief in the same way that bar code describes frame A. However, Dennett observes that frame A can also be described as *bar code with 25% noise* (and so forth) and similarly, there may be more (or less) detailed descriptions of Hanako's belief. If John's Toyota is green and Hanako knows that, then "John drives a green Toyota" may be an alternative description of the same belief (but whether it is depends on what Hanako exactly believes, of course).

Dennett wonders whether *bar code* is really there in frame A. Analogously, the question here is whether the proposition "John drives a Toyota" is really there in Hanako's belief, or in Hanako's mind when she believes that John drives a Toyota. A question whether a proposition or propositional description "is really there" sounds odd, however, partially because it conflicts with how we normally talk about propositions and beliefs, and partially because of ambiguities in the notion of belief.

There is a common distinction between dispositional and occurrent beliefs, where the former are beliefs one holds in the back of one's mind and that aren't presently conscious, while the second are conscious thoughts. Some philosophers, such as Tim Crane (2001), have argued that there are no occurrent beliefs, however, and according to Peter Carruthers (2011; 2013) conscious thoughts don't fit the functional profile of beliefs. Nevertheless, even if there are no occurrent beliefs in a strict sense of "belief", if Hanako has an occurrent *thought* that John drives a Toyota and she (dispositionally) believes in the truth of that thought, that is close enough.

Secondly, dispositional beliefs can be explicit or implicit (Dennett 1978). Explicit beliefs are beliefs one already, knowingly has, while implicit beliefs are beliefs that one would form in the right circumstances given the explicit beliefs one already has. If Hanako's belief is an explicit belief and she knows that Toyotas are Japanese cars, then she implicitly believes that John drives a Japanese car. (She'd probably also implicitly believe that John's car isn't made of paper, for example, as well as very many other things.)

If there are alternative descriptions of one and the same belief as suggested by Dennett's analogy, then there is no clear boundary between implicit beliefs and alternative descriptions of the same belief. If Hanako believes that John drives a green Toyota, then "John drives a Toyota" could either be an alternative (less specific) description of the same belief, or it could be a description of an implicit, different belief. While in this example it seems quite plausible that these are two descriptions of one and the same belief, it is easy to come up with cases in which that is far less clear. (Does "John drive a car" describe the same belief?) By definition, an implicit belief is a belief that a person doesn't *explicitly* hold already, so if some proposition q describes a belief that can also be described by p, and $p \rightarrow q$, then, by virtue of the fact that q describes a belief that the subject already explicitly holds, q cannot describe an implicit belief; but the other way around, that q is not a description of an implicit belief is insufficient to conclude that q and p describe the same belief. (On the problem of distinguishing implicit beliefs from alternative descriptions of the same belief, see also section 3.4.)

For now, we'll ignore implicit beliefs and assume that Hanako's belief that John drives a Toyota is either an occurrent thought or an explicit, dispositional belief. In either case, there are two sources of noise — physical and mental — but there are some differences in their role and nature.

Physical "noise" is the least interesting of the two. Features of Hanako's brain architecture (which is partly dependent on her experiences), but also mind-altering drugs, electronic brain stimulation, and aspects of the physical environment create the "noisy" mental environment (or brain states) in which the thought that John drives a Toyota occurs, but Hanako's brain architecture and other physical/environmental factors also affect a dispositional belief with similar content.

Mental "noise" comes from other beliefs and from what David Kaplan (1968) called the "genetic character" of their components (see section 1). If Hanako has an occurrent thought that John drives a Toyota, then that thought occurs in a context that is constituted primarily by other beliefs (and to a lesser extent by the aforementioned physical factors). This includes preceding and simultaneous thoughts, perceptual beliefs, and related dispositional beliefs. Some of these may be conscious, but many will be unconscious. And all of these adjacent, overlapping, and co-occurring beliefs affect the content of the thought that John drives a Toyota. To translate this into the patterns analogy: frame A is preceded by and co-occurs with other frames (that are or can be described by other patterns) and these various frames infect and bleed into each other.

The most important kind of mental "noise" is due to the interrelatedness or interdependence of many of our (or Hanako's) beliefs. Donald Davidson (e.g. 2001[1997]) famously argued for a kind of holism that ultimately connects all of our beliefs, but we don't have to go that far (although I think Davidson was right in this respect): it is sufficient to realize that "John" is not some kind of abstract or empty marker in Hanako's mind, but refers to a person she knows, a person she has shared experiences with, a person she has memories about.

Hanako has very many other beliefs about John — that he is tall, that they had lunch together last Monday, that he has a cat, and so forth — and importantly, she doesn't have a concept of JOHN *separate* from those beliefs. Rather, these beliefs together make up *her* concept (or mental file or dossier — see section 1) of John, in the same way that my beliefs about tables make up *my* concept of TABLE. (Or in other words, there is no *John* pattern that is later disrupted by noise, but rather, Hanako's concept of John emerges from the noise.) That doesn't mean that all of those beliefs are of equal (John-defining) importance (or equally salient) in all circumstances, however. In the contrary, thinking about John's car most likely activates (consciously or unconsciously) different stored beliefs about John than thinking about last time she had lunch with him, and consequently, the John in Hanako's beliefs is a context-dependent and ever-changing weighted subset of her beliefs about John.

Moreover, the same is true for the other notions involved in the belief that John drives a Toyota. The concept of DRIVING A CAR, Hanako's mental representations of TOYOTA and JOHN'S TOYOTA, and the idea of JOHN DRIVING A CAR aren't abstract or empty markers either, but are similarly constituted by the beliefs in which they

figure.

The implication is rather obvious: Hanako doesn't *just* believe that John drives a Toyota, she believes that some specific John drives (in some specific way) some specific Toyota. It is a green Toyota with a long scratch on the left passenger door, and with stains on the floor from the time when Hanako threw up in his car after his driving style made her sick, and so forth. Much of this remains unconscious, of course, but it is this kind of noise that determines what Hanako exactly believes. Even if Hanako has an occurrent thought "John drives a Toyota", it is inescapably *that* John, and *that* Toyota.

"Mental contents are more specific than the 'that'-clauses used to characterize them", wrote Kent Bach (1997, p.235). *That* is what the noise in Dennett's analogy represents: the inherent specificity of our mental contents; all the idiosyncratic details in the mental representations that give content to our beliefs. And this noise is not a disruption of an originally pure signal. Rather, this noise *constitutes* the "signal".

2.2. "Noisy" mental content

A few years ago, at a small conference involving philosophers and neuroscientists, one of the neuroscientists in attendance claimed (during lunch break) that given enough time with a subject, he would be able to tell what that subject is thinking just by means of neuroimaging — he would be able to "read" mental contents from an fMRI scan. There was a catch, however: with a different subject, he'd (almost) have to start over again.

I don't know whether this neuroscientist was bragging — provided that the mental contents he expected to identify are not excessively fine-grained, his claim doesn't seem implausible to me — but it doesn't really matter, because it's the "catch" that is most interesting.

Let us re-apply Dennett's pattern analogy. This time, think of the six frames in figure 1 as fMRI scans or some other kinds of images of brain activity, and think of the pattern *bar code* as a kind of mental content or mental event — remembering the smell of freshly cut grass, for example. The "noise" that to a greater or lesser extent "obscures" the pattern/mental event in the six frames/fMRI scans is the noise we encountered above: it is due to subtle differences in brain architecture, to preceding and co-occurring mental activity, and most importantly, to different memories, different desires, different beliefs, and so forth. When one person remembers the smell of freshly cut grass, she may remember (or unconsciously activate a memory of) some particular event that involved freshly cut grass, but another person will have different memories and different associations, and will, therefore, activate (however subtly) different networks of mental contents and their neural correlates.

While remembering as a general mental activity will involve very similar parts of

the brain in most people,[4] the details of the content of the memory, and whatever is linked to that mental content through association and (implicitly, unconsciously) activated related memories differs from person to person. And these differences can be so vast that it makes no sense to speak of a pattern corresponding to "freshly cut grass", except in one brain, and in one short time period (because new experiences will change an individual's memories and mental associations). When it comes to specific mental contents rather than broad kinds of mental activities, frame F (in figure 1) may be the closest analogy. In other words, while it is possible to link types of mental activity (like remembering) to types of brain activation patterns — the hippocampus appears to play a key role in this kind of memory, for example — there is no type/type relation on the level of mental contents. There is, of course, a specific brain activation pattern that corresponds with one person's thought of freshly cut grass at one particular time and in one particular set of circumstances, but that is token identity, not type identity.

When it comes to *what* (rather than *that*) people are thinking, feeling, remembering, desiring, and so forth, Dennett's analogy implies a rejection of type identity between the mental and the neuro-physical, while assuming token identity. (It doesn't in any way *prove* or imply token identity, but giving up token identity in addition to type identity would completely disconnect the mind from the brain, leading to some untenable kind of substance dualism.) Probably, the best known theory of mind and mental content based on token identity is Donald Davidson's anomalous monism (2001 [1970]; 2005 [1993]; 2005 [1995]), but there are fundamental differences between that theory and what I am suggesting here. Davidson's anomalous monism is an *a priori* argument that infers monism from the assumption of anomalism. Here, I am assuming (physicalist) monism (on more or less Quinean naturalistic grounds) and using Dennett's analogy to argue against type identity (but not for anomalism).[5]

Davidson's anomalism holds that there are no strict (causal) laws that connect mental to other (non-mental and mental) events. A strict law states that a certain kind of event in certain conditions is sufficient for the occurrence of another kind of event. Laws in physics and chemistry, for example, are strict, but mental "laws" — according to Davidson — are not. Thus, it is a strict law that heating water to 100 degrees Celsius will cause it to boil, but it is not a strict law that thinking of *umeboshi* (Japanese dried, pickled plum) will cause hypersalivation. Supposedly, the second is not a strict law because there are many conditions that could prevent someone from (almost) drooling while thinking of umeboshi. The subject may suffer from some medical condition, for example. Someone unfamiliar with umeboshi would, of course, also not experience hypersalivation, but that is because such a person cannot *think* of umeboshi. However, there also are conditions — such as high pressure — that would prevent water from boiling at $100°C$. One could, of course, add some kind of *ceteris paribus* clause to the water/boiling law (Davidson 2005 [1993]), but the

exact same option is available in case of the umeboshi/drooling "pseudo-law". And while there *might* be more different kinds of conditions that prevent hypersalivation than boiling, that would merely be a difference in degree, and would be insufficient for the *strict* distinction between "strict laws" and mere generalizations that Davidson requires.

My aim here, is not to criticize Davidson's anomalous monism, however, but to contrast it to the noisy mental contents picture emerging from Dennett's pattern analogy. For the latter it doesn't matter whether we call the umeboshi case a "law" or a "generalization". What does matter is that the thought of umeboshi will correspond to different — but probably not *completely* dissimilar — brain activation patterns in different subjects because these different subjects have different histories of encounters with umeboshi, and therefore, different mental representations of umeboshi. (And hypersalivation may trigger further memories and associations that also differ from person to person.) Or, as Davidson once put it, "the correct interpretation of what a speaker means is not determined solely by what is in the head; it depends also on the natural history of what is in the head" (2001[1988], p.44).[6]

2.3. Conceptual representations in the brain

While an analogy may be illustrative and even illuminating, it doesn't prove anything. The noisy mental contents picture emerging in the preceding sections is also supported by a growing body of empirical evidence, however. In a review of research on conceptual representations in the brain, Markus Kiefer and Friedemann Pulvermüller (2012) conclude that concepts are modality-specific, distributed representations that are grounded in perception and action and that are partially coded in the sensory and motor areas of the human brain. Mental representations depend on experiences with the referent(s) of the concepts (Davidson's "natural history" or Kaplan's "genetic character"), which create conceptual memory traces in modality-specific parts of the brain. And concepts are flexible and situation-dependent: "they are comprised of semantic features which are flexibly recruited from distributed, yet localized semantic maps in modality-specific brain regions depending on contextual constraints" (p.817). "To summarize, converging evidence indicates that concepts are flexible, experience-dependent modality-specific representations distributed across sensory and motor systems" (p.817). "Noise" as a metaphor applied to mental content referred to exactly this kind of experience-dependence, situation-dependence, memory-dependence, and so forth (see preceding sections).

"Noise" is the consequence of, among others, different experiences and differences in memory traces and how they are processed. The brain is not a computer hard disk or video recorder and memory is not simple retrieval or replay. Rather, remembering is a constructive process (Schacter, Norman and Koutstaal 1998; Schac-

ter, Guerin and St Jacques 2011; Schacter 2012); "remembering involves the on-
line reconstruction of memorial representations from incomplete and disaggregated
episodic information stored in patterns of neural activation across disperse cortical
areas" (De Brigard 2014, p.410). Furthermore, not just our memories of concrete
objects, events and named particulars are "noisy" in this sense, but *all* our concepts
and other mental representations are. Words and concepts are not (stored as) some
kind of empty abstractions, but activate perceptual and/or motor representations
(Zwaan and Kaschak 2009) and "are typically understood against background situa-
tions" (Barsalou 2009, p.253). And this is not just true of proper names and words we
use for everyday objects, but even of abstract concepts — those too trigger memories
or imaginings of concrete situations (Barsalou and Wierner-Hastings 2005).

All these idiosyncratic details cause "noise", but rather than obscuring our mental
contents, this noise constitutes them. If it wasn't for our experiences and memories
of things, we wouldn't have concepts or mental representations of those things at
all. Nevertheless, even though our mental contents are inherently noisy, we can only
describe them by ignoring much of that noise. No verbal description can do justice to
all of the noisy details, all the memories, all the associations, and all other idiosyn-
crasies that together make up my mental representation of APPLE, your concept of
PEAR, or Hanako's concept of JOHN'S TOYOTA.

2.4. Analyzing belief contents

Thus far the focus has been on (the noisiness of) mental representations or concepts.
Assuming that Hanako's beliefs that John drives a Toyota and that John owns a cat
are — in her mind — about the same John, her concept of John must somehow be
part of or involved in both beliefs, and the same applies to other concepts involved
in her beliefs. Concepts, then, are the recurring building blocks of beliefs. But how
exactly a belief is constructed out of these building blocks is less clear.

The simplest analysis of the content of Hanako's belief that John drives a Toyota
would be that it is an ordered set consisting of her mental representations of JOHN,
DRIVING, and TOYOTA, but that cannot be right. Hanako believes that John drives
some specific Toyota and in some specific way. (It is a green Toyota with a long scratch
on the left passenger door, and with stains on the floor from the time when Hanako
threw up in his car after his driving style made her sick. See section 2.1.) Hence, her
belief doesn't involve a generic concept TOYOTA (if she has such a generic concept
at all), and neither does it involve a generic concept of DRIVING. It seems an obvious
solution for this problem to substitute JOHN'S DRIVING and JOHN'S TOYOTA for
DRIVING and TOYOTA, respectively, but that won't work either. Perhaps, John also
owns an antique Peugeot, which he drives very carefully (contrary to the way he
drives his Toyota). Hence, the concept of DRIVING involved isn't just specific to John,

but also to John's Toyota. But even that won't do, as John may own two Toyotas, one old and green one, which he drives like a maniac, and one brand new, red one, which he drives very conservatively. (And even if he would drive both Toyotas in the same way, JOHN'S TOYOTA would still be insufficiently specific.)

Regardless of how many modifiers are added, the problem doesn't go away. JOHN'S GREEN TOYOTA won't do, because he might have two green Toyotas. And so forth. What must be kept in mind, however, is that the words in small capitals are nothing but labels for Hanako's noisy and idiosyncratic mental representations. Some labels may be badly chosen, others may be better, but ultimately all of them are just labels, and rarely will a label be able to do justice to all of the "noise" in whatever it is labeling. However, a good label should somehow remind us of the inherent noisiness of what it is labeling, rather than obscuring it. We could, for example, use "*THAT* TOYOTA" to refer to Hanako's mental representation of John's Toyota. The deictic "*THAT*" seems an appropriate way of capturing the unavoidable incompleteness and context-dependence of the labels used, but prefixing all labels for mental representations with "*THAT*" also feels rather uncomfortable. A reconstruction of the content of Hanako's belief as ⟨*THAT* JOHN, *THAT* DRIVING, *THAT* TOYOTA⟩ looks odd more than appropriate, even if it is an improvement over ⟨JOHN, DRIVING, TOYOTA⟩. A better option would be to use some neutral symbol as a reminder of the fact that these labels are mere labels and not specifications of mental content — an asterisk will do fine, for example. Then, the content of Hanako's belief is represented as ⟨*JOHN, *DRIVING, *TOYOTA⟩.

The asterisk marks noisiness, but also one of its main implications: uniqueness. If Hanako's thought of a Toyota activates different memories, different associations, and different neural networks than Andrea's thought of a Toyota, then, even if both believe that John drives a Toyota, they really believe quite different things. Moreover, Hanako cannot even have the same belief at two different points in time. Because of new experiences (including new thoughts) the memories, associations, and neural networks involved in Hanako's mental representation *TOYOTA change,[7] and therefore, if Hanako last year believed that John drives a Toyota, and right now still/also/again believes that John drives a Toyota, then Hanako-now and Hanako-last-year really believe(d) quite different things. Beliefs are very much like Heraclitus's metaphorical river: it is as impossible to have the same belief twice as it is to step into the same river twice.

However, if belief reports merely *describe* beliefs, then, even if Hanako and Andrea (or Hanako-now and Hanako-last-year) cannot have the *same* belief, this doesn't imply that we can never use the same *description* of their beliefs in the same context. It may be perfectly acceptable in some (probably even many) circumstances to report that both Hanako and Andrea believe that John drives a Toyota, and therefore — in a sense — share a belief. (See also section 3.1.)

3. Describing beliefs

In the block quote in section 1 Daniel Dennett observes that there are "different attitudes we take at various times toward patterns" (p.35). How we describe frame A depends on circumstances, purposes, technical limitations, and so forth. The same applies to how we describe Hanako's belief. "*Sometimes we care about exact description or reproduction of detail, at whatever cost*". What Hanako believes, then, is that some specific John drives (in some specific way) some specific Toyota, and all of this needs to be specified in detail. (That this may be impossible in practice is irrelevant here.) "*At other times we care about the noise, but not where in particular it occurs*". Again, Hanako believes that some specific John drives (in some specific way) some specific Toyota, but this time we omit the specifications, recognizing the noise, but not describing it in detail. "*And sometimes, we simply tolerate or ignore the noise*". Then, we just say that Hanako believes that John drives a Toyota. This, of course, is what we do most of the time, and in many circumstances that is good enough.

As mentioned before, the key question for Dennett is whether *bar code* is really there in frame A, but the focus here is on truthful (or faithful) description rather than on ontology.[8] There are "different attitudes we take at various times toward patterns", but when is which attitude right or appropriate? When we say that *x* believes that *p* we are ignoring the noise (i.e. we're taking the third of Dennett's three attitudes), but when are we justified to do so, and/or to what extent?

Dennett's pattern, *bar code*, describes frame A in figure 1. Similarly, Mark Richard (1990) and Kent Bach (1997) argue that belief reports *describe* or *represent* what someone believes. "In an attitude ascription we use a sentence to represent what someone think*s*", writes Richard (p.265), and Bach points out that a belief report "does not specify what the person believes but merely describes it" (p.215). And as in Dennett's analogy, such descriptions can be true, appropriate, faithful, *or not* depending on context. Richard suggests that "what counts as faithful representation varies from context to context with our interests and expectations" (p.3), and according to Bach, belief reports are "not true or false independently of context" (p.238).

Strictly speaking, to say that some description or expression is "appropriate" (or "faithful") is a pragmatic claim, while to say that it is "true" is a semantic claim. I will not distinguish these two kinds of claims in the following, however. If in a certain context we *should* care about the "noise" but don't, then the resulting belief description is false and inappropriate in that context, and "falsehood" and "inappropriateness" *here* mean exactly the same thing. This may seem a sloppy use of terms, but it is important to realize that under a stricter interpretation of "falsehood" (that is, stricter than one that effectively identifies it with inappropriateness) *all* belief descriptions — except those that specify *all* noise (but that is impossible) — are false.

Furthermore, the idea that the truth of an imprecise expression depends on the

context in which that expression is uttered is less exotic than it might seem at first. "The distance between Paris and Stuttgart is 500km" is true or false depending on how much precision we require, where exactly we locate the two reference points, whether we are talking about Euclidean, geodesic, or travel distance, and in case of the latter, by what means of transportation. If the context is car travel, then the statement is false (because the shortest distance between the two cities by car, depending on exact starting point and finish, is approximately 620km). If, on the other hand, the statement is made in a discussion about fuel requirements for airplanes, then it is true. There is no answer to the question "What is the distance between Paris and Stuttgart?" that is true in every context. Even if we'd agree on the type of distance and on the two reference points, different contexts may require different degrees of precision.

Stuttgart is called Schduagert in Swabian (a regional language or dialect spoken in Baden-Württemberg and its capital, Stuttgart), and it's not too difficult to come up with a scenario in which someone — let's call him "Hans" — believes that the distance between Paris and Stuttgart is 500km and the distance between Paris and Schduagert is 620km (because Hans doesn't realize that Stuttgart is Schduagert). This is, of course, a variant of Frege's puzzle (see section 1). Because Stuttgart is Schduagert, semantic innocence suggests that "Stuttgart is 620km from Paris" is an alternative description of Hans's belief that Schduagert is 620km from Paris, leading to the attribution of contradictory beliefs.

In case of the more or less "noisy" beliefs implied by Dennett's pattern analogy, there is an asymmetrical relation between different descriptions of the same belief. "John drives a Toyota" is implied by "John drives a green Toyota", but not the other way around. Alternative descriptions of the same belief in variants of Frege's puzzle are materially equivalent, however. "Stuttgart is 620km from Paris" implies "Schduagert is 620km from Paris" and *vice versa*. A second important difference between the two kinds of cases is that while Hanako will almost certainly realize that "John drives a green Toyota" implies "John drives a Toyota", Hans will *not* realize the equivalence of the two alternative descriptions of his belief about the distance between Stuttgart and Paris. In this section, the focus will be on more or less "noisy" descriptions of Hanako's belief; we'll turn to Frege's puzzle and similar cases in section 4.

3.1. Beliefs and descriptions

Traditionally, a belief report is assumed to be a relation between a believer and a proposition, which is formally represented by means of a two-place predicate B. Thus, "$B(x, p)$" means that x believes that p. Dennett's pattern analogy shows that this is misleading, however. That is, p is not what x believes, but merely describes it.[9] Hence, x has a belief b, which is described by p, and — as suggested by the pattern

analogy — that description may be true (or appropriate, or faithful) depending on context c. Rather than "$B(x,p)$", then, we should write "$\exists b[D(x,b) \wedge S(b,c,p)]$" — x has a belief b, and that belief b is appropriately described as p in (given) context c.[10] Applied to the case of Hanako and her belief that John drives a Toyota, the traditional representation of a belief report (or attitude ascription) is "$B(h,t)$" in which h stands for Hanako and t for "John drives a Toyota". What I'm proposing is that it should be "$\exists b[D(h,b) \wedge S(b,c,t)]$" ("Hanako has a belief that is correctly described as 'John drives a Toyota' in the given context") instead.[11]

A few things should be noted about this notation. Firstly, it preserves the dyadic nature of the belief relation. However, rather than assuming that believing is a relation between a believer and a proposition, it takes believing to be a relation between a believer and a noisy mental content. Context is, therefore, not a third argument in believing, but in belief description.

Secondly, if that third argument, context c, is a universally quantified-over variable (rather than a constant) and thus irrelevant, then the two notations are effectively equivalent:

$$\forall x,p[B(x,p) \leftrightarrow \forall c \exists b[D(x,b) \wedge S(b,c,p)]]$$

Such context-irrelevance (or constancy across contexts) may apply to reports of beliefs that cannot be (significantly) affected by noise, such as mathematical and logical truths and, perhaps, other purely abstract beliefs. Hence, if is is true in one context that Luis believes that one plus one is two, then it is probably true in all contexts.

Thirdly, p in "$S(b,c,p)$" *describes* belief b, which suggests that p is a sentence rather than a proposition. However, while "John drives a Toyota" and "John fährt einen Toyota" are different sentences, descriptions of Hanako's belief by means of either sentence are true and false in the same contexts (even though they may not be *understood* in the same contexts), and consequently — at least in this respect — p in "$S(b,c,p)$" is more like a proposition than like a sentence. If p is a proposition, then a more technically correct reading of the S predicate would be something like "belief b is appropriately described in (given) context c by a sentence expressing p". In this paper I will (mostly) take a shortcut and just write that proposition p describes belief b, however.

Fourthly, this notation allows us to solve the problem of shared beliefs mentioned in section 2.4. Recall that Hanako and Andrea (or Hanako-now and Hanako-last-year) have different mental representations of John and his Toyota and thus cannot, strictly speaking, have the same belief that John drives a Toyota. Nevertheless, because "John drives a Toyota" is merely a description of their beliefs (rather than a specification of their belief contents) there may be a context in which that description correctly

describes both beliefs. To say that Hanako and Andrea believe the same thing is really nothing but a loose way of saying that (in the given context) the same sentence accurately describes what they believe(d). Thus, in this loose sense, x and y share a belief that p in context c iff:

$$\exists b_1 b_2 [D(x, b_1) \wedge D(y, b_2) \wedge S(b_1, c, p) \wedge S(b_2, c, p)]$$

and that's all. The two beliefs b_1 and b_2 are not identical — because it is virtually impossible for two beliefs to be strictly identical (see sections 2.2 and 2.3) — but that doesn't matter; all that matters is that after ignoring the contextually irrelevant noise the pattern is sufficiently similar to be described in the same way

Fifthly and finally, the separation of description of belief content from attribution and introduction of beliefs as entities by means of an existential quantifier forces attention on questions about the individuation and identification of beliefs (see also section 2.1), especially if one takes Quine's famous dictum "no entity without identity" seriously. Traditionally, beliefs are identified by their content, but section 2.4 showed that there is no non-problematic way of doing that — because there are many different ways of describing the content of a belief, and because none of those descriptions can *completely* describe the content of a belief, content description does not fix or identify a belief.

Beliefs are not *necessarily* identified by their contents, however — they may be analogous to events in this respect. As Donald Davidson (2001 [1967]) pointed out, one and the same event can be described in very different terms, and events exist independently from their descriptions. Consequently, those descriptions do not identify events. Rather, events are identified by their causes and effects (2001 [1969]) or by their space-time location (2001 [1985]). Like events, belief contents can be described in different ways and are (more or less) independent from those descriptions, and consequently, content-based descriptions don't identify beliefs. Instead, beliefs are or could be identified by the cognitive or functional roles they play (i.e. their causes, effects, associations, overlaps, and so forth), or perhaps — in the case of occurrent thoughts — by their space-time location. However, of these two supposed alternatives, the second, space-time location, is not applicable to dispositional beliefs, and the first, functionalism, may not really be an alternative identification criterion at all. If beliefs play a cognitive role *in virtue* of their content (and/or if cognitive roles cannot be determined without appealing to content), and that seems intuitively plausible, then functionalist identification is not substantially different from content-based identification.

Perhaps, these problems of individuation and identification are really a consequence of demanding more than we need, however. There is no exact boundary between the North Sea and the Atlantic ocean, but few people will deny that the North

Sea exists and most educated people will be able to locate it on a map. (Wittgenstein once asked the similar rhetorical question: "If the border between two countries would be disputed, would it follow that the citizenship of all of their inhabitants would be put in question?" 1967, §556; my translation.) Something similar may apply to beliefs.

The interconnectedness of beliefs mentioned in section 2.1 (as well as some of the other "noise") suggests that there is no meaningful way to strictly separate supposed individual beliefs from total belief states (i.e. from everything a person beliefs),[12] and thus, that there is no exact boundary between Hanako's belief t that John drives a Toyota and (all) her other beliefs. But if the vague boundaries of the North Sea are no reason to deny the existence of the North Sea, then this is no reason to deny the existence of Hanako's belief t either. Furthermore, an inexact description may be sufficient to pick out Hanako's belief and refer to it. Hence, while the suggested representation of the content of Hanako's belief in section 2.4 as ⟨ *JOHN, *DRIVING, *TOYOTA⟩ might have seemed overly sketchy at first, it may very well be the case that it is all we need.

3.2. The closure principle

Hanako's belief can be alternatively described as "John drives a Toyota", "John drives a green Toyota", "John drives a Toyota with a long scratch on the left passenger door", and so forth. As mentioned above, noisier, more detailed descriptions of Hanako's belief entail less noisy descriptions. If p is a noisier description and q a less noisy description then $p \rightarrow q$.

According to Nicholas Rescher (1960), if some person x believes that p and believes that $p \rightarrow q$, then she also believes that q (at least implicitly). This is the widely accepted closure principle of belief — in traditional notation:[13]

CPB.1 $(B(x,p) \wedge B(x,p \rightarrow q)) \rightarrow B(x,q)$

Because the two-place predicate B ignores the noisiness of beliefs and the related context-dependence of belief reports (see previous section), "$\exists b[D(x,b) \wedge S(b,c,p)]$" needs to be substituted for "$B(x,p)$", and so forth, which adds a new argument: context c. In case of the closure principle, context is assumed to be irrelevant, and therefore, must be the same throughout: x has beliefs that are correctly described as p and $p \rightarrow q$ in one context, and therefore, x has another (implicit) belief that is correctly described as q in the same context. Hence, we can revise CPB.1 as follows:

CPB.2 $\forall x, b_1, c, p, q[(D(x,b_1) \wedge S(b_1,c,p) \wedge \exists b_2[D(x,b_2) \wedge S(b_2,c,p \rightarrow q)]) \rightarrow \exists b_3[D(x,b_3) \wedge S(b_3,c,q)]]$

The case of the more or less noisy descriptions is different from that described by the closure principle in three fundamental ways, however. Firstly, p and q do not describe two different beliefs (b_1 and b_3), but are different descriptions of the *same* belief. (In the same way that Dennett's alternative descriptions of frame A are just that: different descriptions, not different frames.) Secondly, it matters here *that* $p \rightarrow q$ and not whether the subject *believes* that $p \rightarrow q$ because there may be contexts in which it is appropriate to say that x believes that q even if x does not believe that $p \rightarrow q$. And thirdly, there is no reason to assume that the context of the consequent is necessarily the same as the context of the antecedent.

This third difference is, of course, the point of Dennett's pattern analogy (as I interpret it here). That one and the same belief can be described in two different ways (p and q) and that one of these descriptions is more general than the other (and thus $p \rightarrow q$) does not imply that they are equally appropriate in the same circumstances. That your neighbor's house burned down because you torched it implies that your neighbor's house burned down, but that does not mean that these two descriptions of the same event are equally appropriate in all circumstances. The same applies to beliefs. That p implies q merely implies that there is a (hypothetical) context in which q describes the same belief, but it does not imply that it is the same context.

Taking these three differences into account, the principle most similar to the closure principle, but applied to one belief under different descriptions, would be:

CPS $\forall x, b, p, q[(D(x, b) \wedge \exists c_1[S(b, c_1, p)] \wedge (p \rightarrow q)) \rightarrow \exists c_2[S(b, c_2, q)]]$

While CPB.2 has three beliefs in one context, CPS has *one* belief in one *or two* contexts. The question about the appropriateness of different "attitudes" we can take towards the pattern(s) and the noise is a question about those contexts. Specifically, it is the question *when* contexts c_1 and c_2 in CPS are non-overlapping such that the belief b cannot be described as q in the *same* context, despite $x's$ belief that p (or, more accurately, despite $x's$ belief described as p) and despite the fact that p implies q. Thus,

NC $\forall x, b, p, q[(D(x, b) \wedge S(b, c, p) \wedge (p \rightarrow q) \wedge \phi) \rightarrow \neg S(b, c, q)]$

wherein ϕ stands for the condition that answers the italicized "when" in the previous sentence.

NC — short for "Non-Closure" — conflicts with CPB.2 if p and q describe the same belief and if $p \rightarrow q$ implies that x believes that $p \rightarrow q$, unless ϕ is impossible. To see this, assume a class of cases in which p and q accurately describe the same belief (but not necessarily in the same context) and in which the antecedent of NC is true. Then, NC implies (obviously) that the belief described as b in context c can *not* be described as q in the same context. However, from CPB.2 it follows that the same

belief b is accurately described as q in the same context (because there is only one context throughout CPB.2).

There are three ways to avoid this contradiction. The most obvious is to assume that ϕ cannot be satisfied — but I see no *a priori* reason to make that assumption. In the contrary — it is fairly easy to come up with an example of condition ϕ, as the following sections will show.

The second is to deny that $p \rightarrow q$ implies that x believes that $p \rightarrow q$. This is technically correct, of course, but if is *obvious* that p implies q then it would be un-charitable to not assume that x believes that $p \rightarrow q$, and if p and q are more and less noisy descriptions of the same belief then it will often be obvious that p implies q. It is, for example, rather implausible that Hanako would *not* realize that "John drives a green Toyota" implies that "John drives a Toyota". Hence, while it isn't *necessarily* the case that $p \rightarrow q$ implies that the believer believes that $p \rightarrow q$ — there may be a context in which it is appropriate to say that Hanako believes that John drives a Japanese car, for example, even if she doesn't know that Toyota's are Japanese cars — in the type of cases considered here x will often (perhaps, even typically) believe that $p \rightarrow q$.

The third, and only remaining option is to exclude cases in which p and q describe the same belief (rather than different beliefs) from the closure principle:

CPB.3 $\forall x, b_1, c_1, p, q[(D(x, b_1) \wedge S(b_1, c_1, p) \wedge \exists b_2[D(x, b_2) \wedge S(b_2, c_1, p \rightarrow q)] \wedge$
$\neg \exists c_2[S(b_1, c_2, q)]) \rightarrow \exists b_3[D(x, b_3) \wedge S(b_3, c_1, q)]]$

3.3. The principle of (avoiding) deviant implications

Normally, what is true of something under a less detailed description holds true under a more detailed description as well. What is true of a table is also true of a wooden table, but not necessarily the other way around, except that "possibly not being made of wood" is true of the first, but not of the second.[14] The same applies to beliefs. If something non-trivial is true of a belief described as "John drives a Toyota" that is not true if the same belief is described as "John drives a green Toyota" (and I take analogues of "possibly not being made of wood" to fall in the "trivial" category), then something went wrong. But the only thing that can have gone wrong is belief description, and consequently, (at least) one of the two descriptions is false. Since it cannot be the more detailed description that is false (for this reason alone), we should in such cases reject the less detailed (or less "noisy"), more general description. Let's call this *the principle of (avoiding) deviant implications*:

> If, in a given context, there is a non-trivial property a belief has under a less specific description of that belief and not under a more specific description, then the less specific description is false in that context.

Formally:

> PDI $\forall x, b, p, q[(D(x,b) \land S(b,c,p) \land (p \rightarrow q) \land \neg(q \rightarrow p) \land \exists F \neg[F(b,p) \leftrightarrow F(b,q)]) \rightarrow \neg S(b,c,q)]$

wherein F is a (contextually) non-trivial property that b has under one description and not under the other. (On the condition "$\neg(q \rightarrow p)$", see section 4.)

Since the last conjuncts in the antecedent, "$\neg(q \rightarrow p)$" and "$\exists \neg F[\dots]$", fill the ϕ slot in NC, an instantiation of PDI is an instantiation of NC. However, I doubt that a property F (that b has or does not have depending on description) is encountered without first assuming both descriptions, and consequently, PDI seems most likely to be found in an argument by contradiction (regardless of whether the argument in question was originally intended as such). That is, if the assumption that q is an accurate description of a belief b that can also be described as p when $p \rightarrow q$ implies that $\exists F \neg[F(b,p) \leftrightarrow F(b,q)]$, then — by PDI — that assumption is false (at least in the given context). Or in (other) words, if it is assumed that "John drives a Toyota" correctly describes a belief that can also be described as "John drives a green Toyota", and that assumption has the implication that there is a non-trivial property that that belief has under one description and not under the other, then that assumption is false, and thus "John drives a Toyota" is *not* an appropriate description of that belief in that context. The next section will illustrate this.

3.4. Gettier cases

Hanako believes that John drives a green Toyota. She has this belief because he has driven a green Toyota for years (the one with the scratch on the passenger door), and she has no reason to believe that this has suddenly changed. She is, therefore, also justified to believe that John drives a green Toyota. Green Toyotas are Toyotas, and therefore, Hanako believes that John drives a Toyota, and for the same reason she is justified to believe that.

However, yesterday, unbeknownst to Hanako, John sold his green Toyota and bought a white Toyota. A white Toyota is still a Toyota, so it is still true that John drives a Toyota.

Hanako believes that John drives a Toyota. She is justified to believe that John drives a Toyota. And it is true that John drives a Toyota. Therefore, Hanako *knows* (by the definition of knowledge as justified true belief) that John drives a Toyota.

This is, of course, a Gettier case — that is, a supposed counterexample against the definition of knowledge as justified true belief (Gettier 1963). Some, but not all Gettier cases fit the same pattern as this example:[15] x has a belief that can be described as p and as q such that $p \rightarrow q$; if it is described as q it counts as knowledge, but — and that always remains unmentioned — if it is described as p it does not count

as knowledge (because it is not true under that description). Hence, depending on description, and without a change of context, the same belief either does or does not have a non-trivial property, namely: "being knowledge" or "counting as knowledge".

Let's quickly go through the steps of the argument to see how it works (or doesn't work, actually). The argument starts with the assumptions that (1) Hanako has a belief b that in the context C is described as g ("John drives a green Toyota") and this belief is justified; that (2) green Toyotas are Toyotas; that (3) John sold his green Toyota and bought a white Toyota; and (4) that white Toyotas are Toyotas too (and let's ignore what Gongsun Long might have to say about this last assumption).[16] It obviously follows from (3) and (4) that (5) John drives a Toyota. So, these five propositions are the start of the argument.

1)	$D(h,b) \wedge S(b,c,g) \wedge J(h,b,g)$	assumption
2)	$g \rightarrow t$	assumption
3)	$\neg g \wedge w$	assumption
4)	$w \rightarrow t$	assumption
5)	t	3,4, elimination and modus ponens

The only new element in these assumptions is the three-place predicate J. "$J(h,b,g)$" means that h (Hanako) is justified to believe b that g (or that h's belief b that g is justified). If all "noise" is taken into account, is seems likely that no two beliefs are completely identical, which means that the first argument (the believer) isn't strictly necessary, so perhaps that could be omitted, but the other two arguments cannot. Beliefs are justified — not propositions — so "$J(g)$" would make no sense, and "$J(b)$" would imply that justification is completely independent from the description of a belief, or in other words, that a belief described as "John drives a Toyota" has the exact same justification as a belief described as "John drives a green Toyota", and as "John drives a green Toyota with a scratch on the left passenger door", and so forth. This seems implausible, and therefore, justification needs at least three arguments. *At least*, because I think it should really be a four-place predicate with whatever justifies the belief in question — the source of justification — filling the fourth place. This doesn't matter in the present context, however, so we can ignore the fourth argument here, but it matters a lot in case of arguments for skepticism based on the closure principle of justification, for example.

The next step is the assumption that t ("John drives a Toyota") is also an appropriate description of Hanako's belief b. In Gettier cases that follow this pattern this assumption is (implicitly) justified by the original, simple version of the closure principle of belief CPB.1 (and the assumption that Hanako believes that $g \rightarrow t$), but the re-revised version of the closure principle, CPB.3, doesn't apply here. Instead, this is where the proof by contradiction starts.

6)	$S(b, c, t)$	assumption for proof by contradiction
7)	$J(h, b, t)$	1,2,6, elimination and the closure principle
		of justification
8)	$K(h, b, t)$	6,5,7, JTB
9)	$\neg K(h, b, g)$	3, elimination and JTB
10)	$\neg S(b, c, t)$	8,9, PDI
11)	$S(b, c, t) \wedge \neg S(b, c, t)$	6,10, conjunction

Line (11) is a contradiction implying that the assumption in line (6) is false and, therefore, that:

12)	$\neg S(b, c, t)$	6–11, proof by contradiction

There are a few steps in this argument that may need some explanation. (7) and (8) introduce the closure principle of justification and the traditional definition of knowledge as justified true belief (JTB), respectively, as well as a new predicate K for "knows". The closure principle of justification holds that if x is justified to believe p and if $p \rightarrow q$ then x is justified to believe q.[17] Knowledge, like justification, is a three-place predicate here: "$K(h, b, t)$" means that h's belief b that t qualifies as knowledge. (8) follows from the definition of knowledge as JTB,[18] the assumption of belief in (6), the inference of truth in (5), and the inference of justification in (7).

In the Gettier-case interpretation, (6) is not an assumption for indirect proof but follows from CPB.1, as mentioned above, and after (8) the argument changes: another assumption is added, namely the intuition that b is not knowledge, which leads to the conclusion that JTB is false. However, CPB.1 was replaced by CPB.3, which doesn't apply here (because g and t describe the same belief), and therefore, (6) doesn't follow. Instead, (6) must be assumed, and this turns out to be an assumption in a proof by contradiction.

Because the first conjunct of (3) is $\neg g$, JTB implies that h's belief b that g is not knowledge (9). Therefore, b is knowledge if described as t (8) and not knowledge if described as g (9). This means that there is a non-trivial property — namely "being knowledge" — that b has under one description and not under the other, and in that case, PDI implies that the more general description t is false (10), which leads to a contradiction (11). Therefore, (12) in this context it is false to say that Hanako believes that John drives a Toyota (or, Hanako's belief cannot be accurately described as "John drives a Toyota" in the given context). This particular case, then, fails as a counterargument against JTB (but that does *not* imply that all Gettier cases fail).

It should be obvious that the foregoing depends on the identification of the beliefs described as "John drives a green Toyota" and "John drives a Toyota" as one and the same belief. The brief discussion of the problem of individuation and identification of beliefs in section 3.1 concluded that beliefs are best identified by their content, and section 2.4 suggested that the best possible representation of the content of Hanako's

belief is something like ⟨*JOHN, *DRIVING, *TOYOTA⟩ in which the asterisk represents the unavoidable incompleteness and context-dependence of the labels used. These labels — "*JOHN", "*DRIVING", and so forth — are just that: *labels*. They *name* (and thus, refer to) mental representations, but due to "noise" (see section 2), they cannot possibly fully describe them. Only sufficient knowledge about context and about Hanako's beliefs can clarify whether two descriptions of what Hanako believes are descriptions of the same belief or of two different beliefs.

Furthermore, there are no shortcuts. Alternative descriptions of Hanako's belief thus far all have the form "John drives a . . . Toyota", wherein zero or more adjectives and/or descriptive phrases are substituted for ". . .", but this pattern is not a useful way of identifying sentences as being descriptions of the same belief — there almost certainly are circumstances in which "John drives a car" or even "John drives" are acceptable descriptions of the same belief. And similarly "John" can be replaced by a long list of contextually appropriate descriptions. On the other hand, if Hanako infers from her belief that John probably has a driving license, then that is not a plausible description of the same belief — that is a new belief, even if it has close ties to the original belief.[19] "John drives a Japanese car" is a much more ambiguous case. Perhaps, it could be a description of the same belief, or perhaps it really is a different, implicit belief. It would require a lot of knowledge about Hanako and her beliefs to decide which it is. (On the distinction between implicit beliefs and alternative descriptions of the same belief, see also section 2.1. The argument in the last paragraphs of section 3.1 — including Wittgenstein's rhetorical question — also applies here: that there are cases that are hard to classify does not imply that *all* cases are hard to classify.)

The practical problem of deciding whether two propositions describe the same belief or two different beliefs (in particular cases) is only of limited relevance here, however. The case of Hanako and John's Toyota presented in this section illustrates PDI and thus when a more general, less "noisy" description of a belief is false. That it may be more difficult in some other cases to see whether two sentences describe the same belief or two different beliefs does not matter for this illustration.

4. Frege's and Kripke's puzzles

PDI (see section 3.3) does not apply to variants of Frege's puzzle (see section 1). In fact, the condition "¬$(q \rightarrow p)$" in PDI explicitly excludes them. If "Superman is strong" and "Clark Kent is strong" are materially equivalent and both are possible descriptions of Lois Lane's belief b, then "Lois Lane accepts x as an accurate description of her belief b" (or the opposite!) would be an example of $F(b, x)$. Regardless of which description of Lois Lane's belief one would choose as p, the other description would

then be false, and consequently, a version of PDI without "$\neg(q \rightarrow p)$" would lead to paradox. Therefore, PDI cannot apply to variants of Frege's puzzle (but perhaps that should have been obvious as PDI is about more or less "noisy" descriptions and Frege's puzzle cannot be understood in such terms; see section 3). This, of course, means that another criterion is necessary to decide between materially equivalent descriptions of the same belief.

The root of Frege's puzzle is that the names "Superman" and "Clark Kent" refer to the same individual and are, therefore, generally assumed to be intersubstitutable, but there is something very peculiar about this idea, even if it may seem intuitively plausible. Take the sentence "Lois Lane believes that Superman is strong". In that sentence, "Lois Lane" refers to Lois Lane and "Superman" to Superman. Whether and to what "believes" and "is strong" refer is controversial, so I'll ignore those terms here. Importantly, this sentence is a belief report, so it must also somehow refer to the belief it is supposed to report. But then what refers to that belief? The only possible answer is the that-clause. That is, "Superman is strong" refers to Lois Lane's belief.

"Superman is strong" is not the content of Lois Lane's belief — it is merely a description, a tool to pick it out and talk about it (i.e. to refer to it), and it *cannot* be more than that. Belief contents are too "noisy" to be pinned down and described *completely* (see sections 2.1 and 2.3). The best we can do (in many, but probably not all cases) is to say that the content of Lois Lane's belief is something like ⟨*SUPERMAN, *STRONG⟩ (see section 2.4). It is not the case, however, that this implies that "Superman" also refers to Lois Lane's mental representation *SUPERMAN — rather, "Superman is strong" refers to Lois Lane's belief as a whole.

What is peculiar about the assumption of intersubstitutionality is that it ignores reference to the belief — even though the context is a belief report. A substitution of "Clark Kent" for "Superman" does not just change the name used to refer to one and the same individual, but also changes the expression used to refer to Lois Lane's belief, and that change cannot be ignored. If "Clark Kent is strong" is not an accurate description of Lois Lane's belief, then this substitution makes the belief report false because it no longer succeeds in referring to Lois Lane's belief. The question, then, is what makes a belief report false (but as mentioned a few paragraphs back, PDI cannot answer this question in case of Frege's puzzle-type cases).

Perhaps, the most obvious answer (which was already suggested above) is that Lois Lane would deny that she believes that Clark Kent is strong. Or in other words, a belief description is false if the believer would not recognize that description as a description of her belief. This answer doesn't work, however, if we want to attribute beliefs to animals. Recall that John has a cat. This cat is normally very fond of John, but if John puts on a long brown coat and a hat, then his cat doesn't recognize him and hides behind the couch. In other words, both "John's cat believes that the man in the brown coat is scary" and "John's cat does not believe that John is scary" are true,

which is, of course, just another variant of Frege's puzzle (or of Quine's Ortcutt case — see section 1). John's cat, however, cannot affirm or reject any description of its beliefs.

Even though Lois Lane's belief that "Superman is strong" refers to her belief ⟨*SUPERMAN, *STRONG⟩ as a whole (see above), there is an obvious — and probably necessary — relation between the components of that belief description and that representation of her mental content. For the belief description to be true, "Superman" must somehow map to Lois Lane's mental representation *SUPERMAN, and so forth. Because it does, "Superman is strong" is a (contextually) correct description of her belief. "Clark Kent is strong", on the other hand, is a false description of her belief, because Lois Lane believes that Clark Kent and Superman are two different individuals and, therefore, "Clark Kent" cannot possibly map to her mental representation *SUPERMAN. Similarly, "John is scary" is a false description of the cat's belief, because "John" cannot possibly map to the cat's mental representation of the man in the brown coat.

The general principle is that a belief report is false if it reports a belief by means of a name or description that refers to x but that cannot map to a mental representation of x in the believer's mind. This may seem to put much weight on the notion of mapping names and descriptions to mental representations, but we don't need to know how exactly such mapping works — all we need is the ability to recognize when a mapping cannot possibly be right; for example, because the name or description obviously maps to a mental representation of what the subject believes to be something or someone else entirely.

In *Naming and Necessity*, Kripke (1980) suggested two further puzzles that seem superficially similar to Frege's puzzle, but that cannot be solved by the principle suggested in the previous paragraph. In Kripke's first puzzle, the monolingual Frenchman Pierre learns about London — known to him as *Londres* — and comes to believe that London is pretty. Some time later, he actually ends up in (a rather dreary part of) London, not knowing it is the same city he knows as *Londres*, learns English, and comes to believe that London is not pretty. In the second puzzle, Peter learns about a musician named "Paderewski" and forms the belief that Paderewski has musical talent. He also learns about a politician of that same name, and thinking that politicians cannot be good musicians, he believes that *that* Paderewski does not have musical talent. He doesn't realize, however, that the two Paderewski's are really the same. Hence, it appears that Pierre and Peter have contradictory beliefs about London and Paderewski, respectively: Pierre believes that London is pretty and not pretty, and Peter believes that Paderewski has musical talent and does not have musical talent.

In these puzzles the same name (or almost the same in case of *Londres*) maps to two different mental representations, and thus the principle solving Frege's puzzle suggested above doesn't work here. But these two puzzles are solved by PDI. Recall

that (almost) any belief has more and less noisy (possible) descriptions (the belief that $1 + 1 = 2$ may be an exception), and that according to PDI a less specific (or less noisy) description of a belief is false if it has "deviant implications" (meaning that the belief has a non-trivial property under one description and not under the other — see section 3.3). In the case of Peter/Paderewski, the more general belief descriptions "Paderewski has musical talent" and "Paderewski does not have musical talent" have the non-trivial property of implying that Peter has contradictory beliefs, while the more specific belief descriptions that "Paderewski-the-musician has musical talent" and "Paderewski-the-politician does not have musical talent" do not have that property, and therefore — by PDI — the more general belief descriptions are false. In the case of Pierre/London there is a similar neglect of noise leading to apparent paradox. More specific belief descriptions — such as references to different parts or faces of London — avoid attributing contradictory beliefs.

Kripke's two puzzles and the use of PDI to solve them differ from the Gettier case discussed in section 3.4 in two important ways, however. Firstly, in the Gettier case both the more and less "noisy" descriptions (i.e. p and q in PDI) are given, but in Kripke's puzzles only the overly general descriptions are given and better, more specific descriptions need to be inferred from the stories. Secondly, and perhaps more importantly, in case of Kripke's puzzles PDI is applied to two beliefs at once (i.e. to Peter's beliefs about a musician and about a politician, and to Pierre's beliefs about London and *Londres*), while in the Gettier case there is only one belief (namely that John drives a certain car). Consequently, this solution of Kripke's puzzles illustrates that the application of PDI isn't necessarily as simple and straightforward as the Gettier case may have suggested.

5. Charity

In Dennett's pattern analogy beliefs are analogous to "frames", more or less "noisy" bands of 90×10 black and white dots (see section 1 and figure 1). Depending on context and needs, we can take different attitudes towards these frames and their description. Sometimes we need lots of detail; in other cases we can "tolerate or ignore the noise" and describe a frame by means of a simple pattern like *bar code*. Similarly, we (can) take different attitudes towards describing and reporting beliefs. Sometimes we need a more detailed description; at other times we can tolerate or ignore the noise. "What counts as faithful representation varies from context to context with our interests and expectations" (Richard 1990, p.3). We can never specify *all* the noise, however — beliefs are too noisy for that.

Beliefs are noisy because they are constructed out of mental representations that are richer, more detailed, and more idiosyncratic than any description could ever

fully express (see sections 2.1 and 2.3; note that some abstract beliefs like the belief that $1 + 1 = 2$ may be exceptions). A report that Hanako believes that John drives a Toyota, for example, ignores that what Hanako really believes is that some specific John drives some specific Toyota in some specific way. Those details may very well be irrelevant in the given context, but that doesn't imply that all details are irrelevant in all contexts. Often we can tolerate or ignore the noise, but sometimes we need to specify some of it.

It is commonly assumed that belief reports specify beliefs. Thus, in the report that Hanako believes that John drives a Toyota, the latter part, "John drives a Toyota", supposedly specifies what Hanako believes. But the inherent noisiness of beliefs implies that this "specification assumption" (Bach 1997) is false. Belief contents cannot be identified with propositions, and belief reports do not specify beliefs. The traditional view exerts an extraordinarily strong pull, however, and is implicitly assumed in the classification of beliefs as "propositional attitudes" — that is, attitudes we take towards particular propositions — but it should not be forgotten that we do not directly observe beliefs. We do not even observe our own beliefs (Carruthers 2011). Rather, "beliefs are *theoretical* entities postulated for the sake of predicting and explaining behavior" (Klausen 2013, p.190). The common idea that belief contents are like propositions — sometimes called "propositionalism" — is not based on observation or evidence, but on introspection and conjecture, and that introspection and conjecture is itself based on an overly simplistic and outdated picture of minds and mental contents. Contrary to that picture, real minds are noisy (see section 2.3).

Belief reports, then, do not specify beliefs. The expression "John drives a Toyota" does not tell us what exactly Hanako's belief is. Rather, it picks out, describes, and refers to what Hanako believes. Kent Bach (1997) defends a very similar view, but there are three important differences between his approach and mine.

Firstly, Bach's argument is based on an analysis of Frege's puzzle and related puzzles about belief reports and is largely *a priori*. Furthermore, the nature of belief contents does not play a central role in Bach's argument. In contrast, my argument starts with an analogy, but after unpacking that analogy, the foundations of the argument are empirical, and belief content takes center stage. The "noisiness" of beliefs and mental representation that refutes propositionalism is largely due to what David Kaplan (1968; see section 1) called the "genetic character" of names and descriptions (see sections 2.1 and 2.2), or to what Donald Davidson called "the natural history of what is in the head" (2001 [1988], p.44), and there is growing empirical evidence that mental contents are "noisy" in this sense (see section 2.3). That is, mental representations are experience-dependent (and thus idiosyncratic), context-dependent, flexible, and modality-specific.

Secondly, Bach stresses that his descriptivism is not a "hidden-indexical theory in disguise" because it maintains the dyadic nature of "believes", which implies that

belief reports do not tacitly refer to context or anything else. In my view, on the other hand, context is an argument, but in a less straightforward way than typically assumed in hidden-indexicalism. A belief report combines an attribution of a belief, which is a psychological (or mental) entity (see section 2.4), with a propositional or sentential description of that belief, but because the same belief can be described in different ways (as illustrated both by Frege's puzzle and Dennett's pattern analogy), attribution and description need to be separated (see section 3.1).[20] Of these two relations, belief *attribution* is dyadic, but the two arguments are a believer and a noisy mental content (i.e. a belief) (rather than a believer and a proposition), and belief *description* is triadic — its arguments are a belief, a propositional or sentential description, and a context. An important implication of this view is that belief reports can be true or false (even if their truth or falsity depends on context), while according to Bach they are semantically incomplete.

Thirdly, and closely related to this last point, there is a difference in focus. Bach's paper is a contribution to a debate about belief ascription that started with Frege's puzzle, and consequently, he focuses on defending his descriptivism *vis-à-vis* competing theories. My focus is less on defending descriptivism (mainly because I don't think it needs a defense) and more on a question that follows from the fact that the same belief can be described in different ways (i.e. more or less "noisy" as in Dennett's analogy, or by means of alternative coreferring names or descriptions in variants of Frege's puzzle), namely: How to choose between competing descriptions? Or: What makes a belief report that appears to be truthful from one perspective or in one context false in another?

Two "principles" were suggested in this paper in an attempt to (at least partially)[21] answer this question (or these questions). The first and most important is the *principle of (avoiding) deviant implications* (PDI; see section 3.3), which holds that if there is a non-trivial property a belief has under one description and not under another then the more general description is false in that context. While it could be argued that any description or representation of a belief that does not violate this principle is right in a given context, in most cases more economical descriptions are preferable. Specifying (much) more detail than necessary is not a virtue. For this reason, if PDI is the only applicable principle, then a more general rule could be that the right description or representation of a belief is the most (or nearly most) general description that does not have "deviant implications" (i.e. does not violate PDI).

The second, nameless principle was suggested in the discussion of Frege's puzzle in section 4. Like PDI, it is based on descriptivism and on the fact that mental representations are "noisy", but while PDI applies to descriptions that are noisy to different degrees (i.e. more or less specific), this principle applies to cases of apparently equivalent belief descriptions differing only in the name or description used to refer to something the reported belief is about. According to this principle, a belief

report is false if it reports a belief by means of a name or description that refers to something but that cannot possibly map to a mental representation of that something in the believer's mind. For example, "Lois Lane believes that Clark Kent is strong" is a false report of Lois Lane' belief that Superman is strong because "Clark Kent" cannot possibly map to Lois Lane's mental representation of Superman (because she doesn't know that Clark Kent is Superman).

Like PDI, this second principle has the purpose of avoiding the attribution of contradictory or otherwise irrational or deviant beliefs, but this suggests that both principles are really just specific versions of the principle of charity. "Serious deviations from fundamental standards of rationality are more apt to be in the eye of the interpreter than in the mind of the interpreted", wrote Donald Davidson (2004 [1986], p.204). Similarly, belief descriptions with deviant (i.e. contradictory, paradoxical, incoherent, or irrational) implications are more apt to be due to a mistake by the interpreter than to accurately describe what is in the mind of the believer.

References

Audi, R. 2011. *Epistemology: A Contemporary Introduction to the Theory of Knowledge*. 3rd Edition. New York: Routledge.

Bach, K. 1997. Do Belief Reports Report Beliefs? *Pacific Philosophical Quarterly* **78**: 215–41.

Barsalou, L. 2009. Situating Concepts. In: Ph. Robbins; M. Aydede (ed.) *The Cambridge Handbook of Situated Cognition*, pp.236–63. Cambridge: Cambridge University Press.

Barsalou, L.; Wiemer-Hastings, K. 2005. Situating Abstract Concepts. In: D. Pecher; R. Zwaan (ed.) *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thought*, pp.129–63. New York: Cambridge University Press.

Brogaard, B. 2007. Attitude Reports: Do You Mind the Gap? *Philosophy Compass* **3**(1): 93–118.

Brons, L. 2016. Putnam and Davidson on Coherence, Truth, and Justification. *The Science of Mind* **54**: 51–70.

Carruthers, P. 2011. *The Opacity of Mind*. Oxford: Oxford University Press.

———. 2013. On Knowing Your Own Beliefs: A Representationalist Account. In: N. Nottelman (ed.) *New Essays on Belief: Constitution, Content, and Structure*, pp.145–65. London: Palgrave MacMillan.

Crane, T. 2001. *Elements of Mind*. Oxford: Oxford University Press.

Crimmins, M.; Perry, J. 1989. The Prince and the Phone Booth: Reporting Puzzling Beliefs. *Journal of Philosophy* **86**: 685–711.

Davidson, D. 2001[1967]. The Logical Form of Action Sentences. In: *Essays on Actions and Events*, pp.105–22. 2nd Edition. Oxford: Oxford University Press.

———. 2001[1969]. The Individuation of Events. In: *Essays on Actions and Events*, pp.163–80. 2nd Edition. Oxford: Oxford University Press.

———. 2001[1970]. Mental Events. In: *Essays on Actions and Events*, pp.207–25. 2nd Edition. Oxford: Oxford University Press.

———. 2001[1983]. A Coherence Theory of Truth and Knowledge. In: *Subjective, Intersubjective, Objective*, pp.137–53. Oxford: Oxford University Press.

———. 2001[1985]. Reply to Quine on Events. In: *Essays on Actions and Events*, pp.305–11. 2nd Edition. Oxford: Oxford University Press.

———. 2001[1988]. The Myth of the Subjective. In: *Subjective, Intersubjective, Objective*, pp.39–52. Oxford: Oxford University Press.

———. 2001[1997]. The Emergence of Thought. In: *Subjective, Intersubjective, Objective*, pp.123–34. Oxford: Oxford University Press.

———. 2004[1986]. Deception and division. In: *Problems of Rationality*, pp.199–212. Oxford: Oxford University Press.

———. 2005[1993]. Thinking Causes. In: *Truth, Language, and History*, pp.185–200. Oxford: Oxford University Press.

———. 2005[1995]. Laws and Cause. In: *Truth, Language, and History*, pp.201–19. Oxford: Oxford University Press.

De Brigard, F. 2014. The Nature of Memory Traces. *Philosophy Compass* **9**(6): 402–14

Dennett, D. 1978. *Brainstorms*. Hassocks: Harvester.

———. 1991. Real Patterns. *The Journal of Philosophy* **88**(1): 27–51.

Frege, G. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* **100**: 25–50.

Gettier, E. 1963. Is Justified True Belief Knowledge? *Analysis* **23**: 121–3.

Kaplan, D. 1968. Quantifying in. *Synthese* **19**: 178–214.

Kiefer, M.; Pulvermüller, F. 2012. Conceptual Representations in Mind and Brain: Theoretical Developments, Current Evidence and Future Directions. *Cortex* **48**: 805–25.

Klausen, S. 2013. Losing Belief, While Keeping up the Attitudes: the Case for Cognitive Phenomenology. In: N. Nottelman (ed.) *New Essays on Belief: Constitution, Content, and Structure*, pp.188–208. London: Palgrave MacMillan.

Kremer, M. 2010. Sense and Reference: The Origins and Development of the Distinction. In: M. Potter; T. Ricketts (ed.) *The Cambridge Companion to Frege*, pp.220–92. Cambridge: Cambridge University Press.

Kripke, S. 1980. *Naming and Necessity*. Cambridge MA: Harvard University Press.

McKay, Th.; Nelson, M. 2010. Propositional Attitude Reports. In: E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Spring 2014 Edition. https://plato.stanford.edu/archives/spr2014/entries/prop-attitude-reports/. Access: 02/13/2019.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge MA: Harvard University Press.

Quine, W. V. O. 1960. *Word and Object*. Cambridge MA: MIT Press.

———. 1964[1948]. On What There Is. In: *From a Logical Point of View*, pp.1–19. Cambridge MA: Harvard University Press).

———. 1976[1956], Quantifiers and Propositional Attitudes. In: *The Ways of Paradox and Other Essays, Revised and Enlarged Edition*, pp.185–96. Cambridge MA: Harvard University Press.

Rescher, N. 1960. The Problem of a Logical Theory of Belief Statements. *Philosophy of Science* **27**(1): 88–95.

Richard, M. 1990. *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. Cambridge: Cambridge University Press.

———. 1997. Propositional Attitudes. In: B. Hale; C. Wright (ed.) *A Companion to the Philosophy of Language*, pp.197–226. Oxford: Blackwell.

Schacter, D. 2012. Adaptive Constructive Processes and the Future of Memory. *American Psychologist* **67**(8): 603–13.

Schacter, D.; Norman, K.; Koutstaal, W. 1998. The Cognitive Neuroscience of Constructive Memory. *Annual Review of Psychology* **49**(1): 289–318.

Schacter, D.; Guerin, S.; St Jacques, P. 2011. Memory Distortion: an Adaptive Perspective. *Trends in Cognitive Sciences* **15**(10): 467–74.

Schwitzgebel, E. 2015. Belief. In: E. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. Summer 2015 Edition. https://plato.stanford.edu/archives/sum2015/entries/belief/. Access: 02/13/2019.

Taschek, W. 2010. On Sense and Reference: A Critical Reception. In: M. Potter; T. Ricketts (ed.) *The Cambridge Companion to Frege*, pp.293–341. Cambridge: Cambridge University Press.

Wittgenstein, L. 1967. *Zettel*. Berkeley: University of California Press.

Zwaan, R.; Kaschak, M. 2009. Language in the Brain, Body, and World. In: Ph. Robbins; M. Aydede (ed.) *The Cambridge Handbook of Situated Cognition*, pp.368–81. Cambridge: Cambridge University Press.

Notes

[1]Semantic innocence implies that a term has the same sense in all contexts, but "Clark Kent" in Lois Lane's belief that Clark Kent is not strong does not have the same sense as "Clark Kent" in the proposition that Clark Kent is Superman (which is true in the universe of Superman, but unknown to Lois Lane).

[2]This definition is a paraphrase of the definition given by Kaplan on p.203, differing mainly in the substitution of symbols consistent with those used here.

[3]Although Daniel Dennett gave me permission to use the original figure, the only scan available was of such low quality that it added another (unintentional) level of noise. For that reason, I redrew frame A (which is referred to below) and recreated the other 5 frames following the procedure described in Dennett's paper (except that I used real random numbers from random.org).

[4]There are significant differences even in this respect. Congenital conditions (such as aphantasia) and brain damage affect which parts of the brain are involved in memory, but thereby also the form those memories take and how they are connected to other mental activity.

[5]That there are no exact correspondences between content-based types of mental events or states and types of brain activation pattern does not in itself imply or require anomalism. Strictly speaking, it doesn't even refute reductionism. It merely makes reduction (of mental events and contents to brain states) impossible in practice, except perhaps — as suggested by the aforementioned neuroscientist — in single subjects and within a short time span.

[6]David Kaplan (1968) referred to this "natural history of what is in the head" as the "genetic character" of a name. See section 1.

[7]This raises questions about the identity over time of mental representations, of course, but that issue is no more complicated or puzzling than identity over time in general, and much of the same arguments and solutions apply.

[8]The analogous ontological question would be whether "John drives a Toyota" is really there (in Hanako's mind) when or if Hanako believes that John drives a Toyota, but that question doesn't seem to make much sense. At least, I have no clue what exactly it means for a proposition to "be there" (in someone's mind).

[9]Recall the quote by Bach in section 1: a belief report "does not specify what the person believes but merely describes it" (1997, p.215).

[10]The "*D*" comes from Greek "*doxa*"; the "*S*" stands for "deScribe". I'm using "*D*" instead of "*B*" to avoid confusion with the traditional dyadic relation *B* between a believer and a proposition.

[11]Kaplan (1968) also separated belief attribution from representation, but in an importantly different way: "$\mathbf{R}(\alpha, y, x) \wedge x\mathbf{B}^\ulcorner\alpha$ is $F^\urcorner$" stands for "a belief of x that y is F if represented as α" (see also section 1). The most important differences are that Kaplan's formalization is designed especially for Frege's puzzle-type cases and is not applicable to beliefs that don't fit the "α is F" pattern (such as the belief that John drives a Toyota), and that it lacks a context argument.

[12]Donald Davidson (2001 [1983]) made a similar point when he wrote that "there is no useful way to count beliefs" (p.138). See also Brons (2016).

[13]I'm not aware of anyone rejecting the closure principle of belief. Robert Nozick (1981) famously rejected the related closure principle of knowledge, but did *not* reject the closure principle of belief. (See p.208.) However, in this paper I argue for a restriction of the closure principle (see CPB.3), and if the closure principle is defined as a universal and exceptionless "law" then such a restriction is effectively a rejection.

[14]Something like this exception was suggested to me by an anonymous referees of this journal. I owe that referee my gratitude for that suggestion.

[15]See, for example, the Gettier case in Audi (2011, pp.248–7). Of the two cases in Gettier's (1963) paper, the first *can* be interpreted as fitting in this pattern, but the second cannot.

[16]The classical Chinese philosopher Gongsun Long famously argued that a white horse is not a horse. (Or perhaps he just claimed that it is or can be permissible to say that a white horse is not a horse. There is considerable debate about the interpretation of Gongsun Long's *White Horse Dialogue*.)

[17]$\forall x, b, p, q[(J(x,b,p) \wedge (p \rightarrow q)) \rightarrow J(x,b,q)]$

[18]$\forall x, b, p[K(x,b,p) \leftrightarrow_{def} (D(x,b) \wedge S(b,c,p) \wedge J(x,b,p) \wedge p)]$

[19]The reason that the above reconstruction of this particular kind of Gettier case doesn't show the failure of Gettier cases in general is that there are many Gettier cases in which the second belief is unambiguously a new belief.

[20]Kaplan (1968) also separated belief attribution from belief representation. See section 1 and note 11 (i.e. the third note in section 3.1).

[21]It is quite possible that further rules or principles need to be added, but I have no clue what those could be.