Dan Sperber (1997) Intuitive and reflective beliefs. *Mind and Language 12*(1), 67-83.

Abstract: Humans have two kinds of beliefs, intuitive beliefs and reflective beliefs. Intuitive beliefs are a most fundamental category of cognition, defined in the architecture of the mind. They are formulated in an intuitive mental lexicon. Humans are also capable of entertaining an indefinite variety of higher-order or "reflective" propositional attitudes, many of which are of a credal sort. Reasons to hold "reflective beliefs" are provided by other beliefs that describe the source o the reflective belief as reliable, or that provide explicit arguments in favour of the reflective belief. The mental lexicon of reflective beliefs includes not only intuitive, but also reflective concepts. [PDF version]

INTUITIVE AND REFLECTIVE BELIEFS

Dan Sperber

I used to be a full-time anthropologist. Anthropologists often make statements of the form: "The So-and-so believe that…" Few of them have bothered to discuss what they mean by "belief" (the most notable exception being Rodney Needham, 1972). Over the years, I have raised several objections to these attributions of beliefs, and have tried to outline a more fine-grained account of the cognitive attitudes involved (Sperber 1975, 1982/1985, 1990, 1994b, 1996). Just as the common term "jade" correspond to two substances, jadeite and nephrite, with similar phenomenal properties but quite different chemical structures, the folk-psychological term "belief", I have argued, corresponds to two psychological categories, similar in some behavioural and epistemological respects, but different in cognitive organisation and role. I call these two psychological categories "intuitive beliefs" and "reflective beliefs". In this paper, I would like to develop the argument in a manner that addresses questions and objections that I have received from psychologists and philosophers of mind.

Intuitive beliefs

When we claim of an organism that it possesses a cognitive system, we attribute to it at least two kinds of representations. The overall function of a cognitive system is to allow the organism to adjust its behaviour to a changing environment. For this, a cognitive system must contain representations of actual states of affairs, a data-base, so to speak. It must also contain representations of behaviours to be engaged in by the organism, representations capable of guiding these behaviours, in other terms, plans. The simplest link between data and plans consists in having the triggering of every plan-guided behaviour conditional on the addition to the data-base of a specific datum. For instance, if a representation of a cat approaching is added to the data-base of a mouse, this activates a flight plan.

Simple cognitive systems have a data-base, i.e. a place in their functional architecture where representations can be stored, and such that any representation stored in thi place is, without restrictions, acted upon as if it were a representation of an actual state of affairs. It is a moot point whether such representations should be called "beliefs" when they occur in the data-base of organisms that lack a public language and the very concept of a belief. Humans, on the other hand, think and talk of some of their mental states as "beliefs". Unlike eliminativists who deny that the category of belief is instantiated at all, I will argue that it is instantiated by objects belonging to different natural types.

Does the human cognitive system contain a data-base too, where beliefs are stored, or are a person's beliefs beliefs merely in virtue of being conceptualised as such by the believer? Suppose that a person's belief that P was so identified by, say, a prefix representing the person's own attitude of belief such as (1):

(1) Bel (P)

Would it be enough that Mary entertained (1) for her to believe that P? Surely not. She could entertain (1) as an object of doubt, desire, fantasy, etc. For her entertaining [Bel (P)] to amount to her believing that P, she would have to treat [Bel (P)] as factual, to believe it too. Needless to say, embedding it under the same prefix "Bel", as in [Be (Bel (P))], would start an infinite regress. So, in the functional architecture of human minds too, there has to be a data-base such that any representation stored in that data-base is treated as a representation of an actual state of affairs, i.e. as a belief.

What makes the data-base a *data*-base, or a *belief box*, to use Stephen Schiffer's phrase, is that the representations it contains, by the very fact of being so located, are freely used as premises in practical and epistemic inferences. Data-base beliefs are "intuitive" in the sense that, in order to hold them as beliefs, we need not reflect – or even be capable of reflecting – on the way we arrived at them or the specific justification we may have for holding them. The presence of a representation in our data-base causes us to treat it as data. Moreover, if we trust our cognitive mechanisms, then the very fact that a representation has been inscribed by these mechanisms in our data-base is *good reason* to treat it as data.

Every representation stored in Mary's data-base is a belief of Mary's, but the converse is not true, and this for two reasons, one generally acknowledged, the other generally ignored. We may be justified in attributing to Mary the belief that there are no kangaroos on Jupiter, and indefinitely many other comparable beliefs, even though they are not represented at all in her mind. For Mary herself, these tacit or virtual beliefs follow directly from the beliefs that are actually and explicitly represented in her mind. Thus, she would immediately agree – and hence she believes, at least in a dispositional sense – that there are no kangaroos on Jupiter. This observation is compatible with the view, generally taken for granted, that every belief of Mary's that *is* represented in her mind is represented *in her data-base*, and is a belief of hers in virtue of being so represented. I want to argue against this view. There are not one but two ways in which propositional attitudes in general, and beliefs in particular, can be inscribed in the mind.

The very fact that humans can entertain desires, suppositions, or fictional representations, which are not freely used as premises in inference, shows that not all their mental representations are stored in the same way and in the same place. One possibility is that there are other "bases" or "boxes" in the functional architecture of the mind. (Such permanent boxes, which each define a basic type of mental representation, should not be confused with the temporary buffers of inference engines, where a number of premises with different cognitive statuses may be brought together for joint processing.) The existence of a desire "box" is a quite plausible. One might also, in principle, but much less plausibly, have a box for suppositions, a box for guesses, a box for deemings, etc. Multiplying hypothetical boxes is not, however, the only way to account for the multiplicity of propositional attitudes.

Reflective attitudes

Humans have the ability to represent representations. I would argue (see Sperber & Wilson 1986/1995; Sperber 1994a ,1994b) that this meta-representational ability is as distinctive of humans, and as important in understanding their behaviour, as is echolocation for bats. Humans have the ability to represent three types of representations: mental representations, public (e.g. linguistic) representations, and abstract representations. Mary might thus have the following three meta-representations in her data-base:

(2) Peter believes that the Earth is flat
(3) John said that the Earth is flat
(4) The hypothesis that the Earth is flat is absurd

Representations embedded in meta-representations which are themselves stored in the data-base are, in a sense, present in the data-base, but they are insulated from other representations in the base by the meta-representational context in which they occur embedded. They are not automatically treated as data. Mary, for instance, has the proposition that the Earth is flat represented at least three times in her mind, but not as a belief of hers. She cannot detach it from any of the contexts in which it occurs and store it directly in her data-base. This embedded representation may play a role in a number of inferences. Mary is in a position to infer for instance that:

(5) John said something that Peter believes
(6) What Peter believes is absurd
(7) Peter believes that the Earth is not spherical

Drawing inferences about a meta-represented representation may involve detaching it provisionally and copying it in the temporary buffer of some inferential device where it will be processed together with other premises. However, the output of such a process must be re-embedded in an appropriate context in order to arrive at a warranted conclusion. For instance, deriving (7) from (2) might be done by provisionally disembedding (8) from (2), processing it together with some premise such as (9), inferring (10), and re-embedding (10) in the context (11):

(8) The Earth is flat
(9) For all x, if x is flat, then x is not spherical
(10) The Earth is not spherical
(11) Peter believes that…

Having certain meta-representations in one's data-base amounts to having a propositional attitude to the representation meta-represented. Consider, for instance:

(12) It is doubtful that the devil exists
(13) The claim that the Earth is flat has been refuted
(14) Everybody knows that real truffles are expensive

Believing (12) amounts to doubting that the devil exists. Believing (13) amounts to disbelieving that the Earth is flat. Believing (14) amounts to being of the opinion that rea truffles are expensive. "Amounts to" in what sense? In the sense that an individual understanding and believing (12), (13), or (14) will, ipso facto, have the corresponding attitudes to the embedded propositions.

Via the meta-representational route, an indefinite variety of different propositional attitudes may be taken to the meta-represented representation. I will call such attitudes "reflective attitudes." I doubt that reflective attitudes fall into sharply bounded, mutually exclusive, well-defined categories. There is, for instance, a continuum of reflective attitudes between absolute conviction and radical disbelief, with all shades of doubt in between.

Reflective beliefs

Among reflective attitudes, a great many are, more or less prototypically, credal attitudes, that is, attitudes of "belief" in the ordinary, somewhat loose sense of the term. (14) provides an illustration, and so do (15)-(20) :

(15) Peter is sincere when he says that he is in pain.
(16) I remember that, the day we first met, it was raining.
(17) There are indubitable signs that someone has been searching the house.
(18) It is a scientific fact that a glass of wine a day is good for the heart
(19) It has been proven that communism does not work.
(20) That the Father, the Son and the Holy Ghost are one is a Holy Mystery.

A person in good faith who believes (14)-(20) should be disposed in each case to assert the embedded representation, or to assent to its assertion by others, and this whether or not the embedded representation happens to occur unembedded in her data-base, i.e. whether or not that representation is a data-base, intuitive belief. The observation of such a person's behaviour would, then, warrant attributing to her the corresponding beliefs, but not necessarily the corresponding intuitive beliefs.

Two rather uncontroversial assumptions about human cognition should lead one to recognise that humans must be capable of holding credal attitudes in at least two ways The two assumptions are:

- The human mind has an ability to hold representations as beliefs
- The human mind has a meta-representational ability

From these two assumptions, it follows that humans are capable of having beliefs about representations. Such meta-representational beliefs may imply (demonstratively o non-demonstratively) that the representation meta-represented is true. In other terms, a belief with meta-representational content may provide a *validating context* for the embedded representation. When this occurs, the individual has, ipso facto, a credal attitude to the representation embedded in the meta-representational belief. In other terms, the individual has two credal attitudes: one with content V(R) where "V" is a validating context, and "R" is a representation; the other with content R. The credal attitude with content V(R) is a data-base belief. The credal attitude towards the embedded representation R is not a data-base belief. I will call such credal attitudes toward representations embedded in a validating context of the form V(R), "reflective beliefs". There is an indefinite variety of possible validating contexts: reference to authority, t divine revelation, explicit argument or proof, etc. Hence, just like reflective attitudes in general, reflective beliefs are best seen as a fuzzy and internally diversified category

Disquotational incontinence

Probably the most common way in which we acquire reflective beliefs is through communication. You ask the railway employee when is the last Sunday train to Oxford. "A 11:45," he answers. You are then likely to form a data-base belief of roughly the following tenor:

(21) The railway employee (who is to be trusted on such matters) said that the last Sunday train to Oxford is at 11:45.

If you hold (21) as a data-base belief, then, by this very fact, you hold (22) as a reflective belief:

(22) The last Sunday train to Oxford is at 11:45.

Such a description raises an obvious objection. If you believe what the railway employee told you, then, surely, you immediately disembed or "disquote" what he said, extract the information from its validating context, and add it to your data-base, where it becomes a regular intuitive belief, which you might keep holding even if you forget how you have arrived at it. In contrasting intuitive beliefs and reflective beliefs, then, I am, it could be objected, contrasting a basic category of genuine beliefs with a mere step in the derivation of some of these beliefs, a step at which their content occurs embedded in a validating context.

Let us call "disquotational incontinence" the thesis that whenever a conceptual representation occurs in a validating context, it is ipso facto disembedded from this context and dropped into the data-base. If humans suffer from disquotational incontinence, then, indeed, it is plausible that all their beliefs are data-base beliefs. However, I want t argue that the disquotational incontinence thesis is false, and that reflective beliefs are a genuine psychological category.

I am not denying, of course, that many of our data-base beliefs have been extracted from validating contexts. For instance, in the (21)-(22) example, you may well have disquoted "the last Sunday train to Oxford is at 11:45" and have stored it in your data-base. Defenders of disquotational incontinence, for their part, would not deny that we can have, in the data-base, beliefs of the form V(R). They would grant, for instance, that you may believe (21), but they would argue that, if you do, this leads you to believe (22), in the same basic sense of "believe", a claim which, in this particular instance, is plausible.

There are two points of disagreement, however:

- Disquotational incontinence means that whenever you hold an intuitive belief with content V(R), you also hold an intuitive belief with content R.
- In order to derive from disquotational incontinence an argument against reflective beliefs, it must be further assumed that when you believe both V(R) and R (as in the case of (21)-(22)), behaviour expressing your belief R is always directly based on your disquoted data-base belief R, and never on a reflective belief R embedded in your data-base belief V(R).

I will argue against these two claims. But first, let me raise an onus of proof issue.

At first blush, it might seem that the onus of proof is on whoever proposes the existence of a new psychological type, in this case "reflective beliefs". However, if you grant me – trivially – that data-base beliefs and a meta-representational ability are part of the basic human psychological equipment, then the possibility of a wide range of propositional attitudes of a reflective kind follows. For instance, if you *believe* that it is doubtful that that the devil exists, then you *doubt* that the devil exists. Doubting is a reflective attitude; it is about representations (e.g. claims, hypotheses), not directly about states of affairs. Most reflective attitudes (e.g. doubting, pondering, disbelieving, accepting as a working hypothesis, granting for the sake of argument, etc.) do not warrant disquotation anyhow. Reflective belief is just one of the many reflective attitudes made possible by the joint existence of data-base beliefs and meta-representations. In order to cast doubt on the existence of reflective beliefs, one must assume a mechanism which systematically disquotes them. To assume disquotational incontinence and deny the actual existence of reflective beliefs is, then, to make an extra hypothesis, and, therefore, to bear the onus of proof.

Defenders of disquotational incontinence might accept the onus of proof and argue that the "proof" requested is ridiculously easy. I will make it even easier by granting that if you have some intuitive belief P, and if you are disposed to infer spontaneously Q from P, then it is reasonable to attribute to you the intuitive belief Q. Well, then: If you have a belief of the form V(R), where V is a context that you understand as validating, then, surely, you are disposed to infer spontaneously R from V(R) – or else, what does it mean to understand that V is validating? Disquotational incontinence follows from trivial assumptions about spontaneous inferability. Right? Wrong!

In earlier papers (Sperber 1982/85, 1990), I presented an argument against disquotational incontinence based on considerations of rationality and good cognitive design. Half-undersood representations such as the dogma of the Holy Trinity can be objects of belief. However, disquoting such half-understood representations and using them unrestrictedly as premises in inference, on a par with well-understood representations, would be, I argued, epistemically hazardous. For instance, contradictions could arise undetected. Half-understood information may be epistemically useful, but only if it is treated with cognitive caution. This argument, which I still find reasonable, is not

quite compelling. It could be, for instance, that our cognitive system is not well-designed in this respect, and indulges in disquotational incontinence in spite of the cost. In this paper, I present a new and I believe stronger argument against disquotational incontinence.

Representational capacity

To accept the existence of a data-base is to assume some representational capacity by means of which data are represented in the data-base. I will, for expository purposes, describe this capacity as a language of thought, a "mentalese," and I will focus on the conceptual repertoire, the lexicon of this mentalese, but the argument should go through, mutatis mutandis, with any kind of account, e.g. connectionist, of the conceptual capacity involved. For the argument to go through, all that is needed is that the representational capacity should have limits, at any given time in the cognitive life of the individual, and surely, that must be true. In a nutshell, I will argue that it is possible to meta-represent representations which are not directly expressible within the conceptual repertoire of the data-base. Such representations can be embedded in a validating context, and yet cannot be disquoted.

Any language, public or mental, has a finite lexicon. In public languages, it is possible to supplement this lexicon with an indefinite range of expressions of the form ["a"], where a may be any symbol, for instance an onomatopoeia, a foreign word, or a word or phrase from the language itself but with a sense different from the one assigned t it by the grammar. Such symbols may be indicated, in writing, by quotation marks, or, in speech, by intonational patterns. However, both in writing and in speech, expressions may be logically quoted without this being explicitly indicated at all. In particular, an expression a should be considered quoted at the logical form level if the grammar of the language does not assign it a semantic interpretation, or assigns it an interpretation quite different from the one with which it occurs in the utterance. Knowledge of the foreign language from which such an expression might be borrowed, or knowledge of special codes, or pragmatic inference, may make it possible to paraphrase the intended meaning, but it may also occur that the intended meaning of the expression mentioned remains more or less obscure.

Mentalese too must allow for a device like quotation marks, and for the use of this device to supplement its lexicon with expressions in quotes. In thinking, we may experience the need for a concept unavailable in our mental lexicon, and which we don't yet grasp well enough to add it to the lexicon. We may meta-represent the yet-to-be-developed concept by means of a mental place-holder. A good place-holder may be a phrase of mentalese in quotes, the meaning of which is evocative of the concept we are seeking.

The need for a quotational device in mentalese is particularly evident in the case of comprehension, and this in more ways than I need to go into here (see Sperber & Wilson 1986/1995). We understand public utterances by associating with them representations in mentalese. An utterance may contain an expression we do not fully comprehend, either because it does not belong to the public language being used and occurs in quotes in the utterance itself, or because our knowledge of the public language is faulty and does not provide us with a meaning for the expression. In such a case, we must be capable of entertaining thoughts containing expressions such as [" a", whatever it may mean] where " a" is a meta-representation of a not fully understood concept.

Imagine young Lisa hears her science teacher say:

(23) There are millions of suns in the universe.

Lisa trusts her teacher, and is willing to believe what he says. However, until that moment, she understood "Sun" as a proper name referring to a single object. She does not know what "sun" might mean as a common noun. She can guess that suns must be things like *the* Sun, and also that the teacher means something more specific than this, and that therefore she does not comprehend the teacher's full meaning. We may therefore attribute to her the following intuitive beliefs:

(24) The teacher (who is to be trusted on such matters) said that there are millions of "suns" in the universe.

(25) There are millions of Sun-like things in the universe

(26) There are millions of "suns", whatever the teacher means by "sun", in the universe.

However we should not attribute to Lisa an *intuitive* belief with content (23), because she does not have the conceptual means of representing it in her data-base. On the other hand, we should not hesitate to attribute to Lisa the belief, in the ordinary folk-psychological sense of "belief", that there are millions of suns in the universe. After all, she is now willing to say (23) herself, and to assent to its expression by others. In so doing, she is not repeating words of which she has no understanding, as she might a Latin formula. She has partial understanding of what she says, as shown by the fact that she is disposed to draw some consequences from it, e.g. that there are many suns.

Young Bob hears his Sunday School teacher say:

(27) The Father, the Son, and the Holy Ghost are one.

Bob trusts his teacher, and is willing to believe what she says. Let us suppose that Bob understands every individual word in this statement. Still, he has only very partial understanding of the statement as a whole. He has only the vaguest and most insecure idea of what the Father, the Son, and the Holy Ghost *being one* might mean. We may, nevertheless, attribute to Bob intuitive beliefs such as the following:

(28) The teacher (who is to be trusted on such matters) said that the Father, the Son, and the Holy Ghost are one.

(29) It must be true that the Father, the Son, and the Holy Ghost are one, whatever this means.

We cannot attribute to Bob an intuitive belief with content (27), since he does not have the conceptual means to represent such a content. The problem here is not so muc that Bob is lacking any concept in particular; it is that his concepts, arranged in such a syntactic structure, don't provide a full-fledged meaning, suggesting that at least one of the concepts expressed by his teacher's utterance is not the one standardly encoded by the words used. On the other hand, Bob is, now, a true believer. He does believe, and will tell you so himself, that the Father, the Son, and the Holy Ghost are one.

Lisa is exhibiting belief-behaviour vis-à-vis (23), and Bob vis-à-vis (27). By folk-psychological standards of belief attribution, they believe what their respective teachers tol them. They could not hold (23) or (27) as intuitive beliefs, but they can, and, I submit, do, hold them as reflective beliefs.

At this point, defenders of disquotational incontinence may want to deny that our folk-psychological standards of belief attribution are strict enough, and demand, for instance, that a belief that P should not be attributed to an individual who does not properly understand P. One problem with such a stricter criterion is that a great many of the beliefs that anthropologists and historians of ideas study would no longer count as beliefs (but, maybe, as "quasi-beliefs", as Recanati 1997 envisages, before himself dismissing the notion). Beliefs in Holy Mysteries such as the Trinity, which are defined as truths beyond the pale of human understanding, would, by definition, not be beliefs at all. Such a terminological move would have obvious costs and no clear benefit (in particular, no explanatory purchase). Anyhow, if there were some smart positiv terminological proposal with clear benefits, I would not particularly want to resist it. My point is not terminological; it is that there are at least two cognitively distinct manner of holding true, i.e. of "believing" as commonly understood.

My argument so far has been that that we can meta-represent more representations than we can construct, and that, therefore we can believe reflectively contents that we cannot hold as intuitively given data. I have focused on half-understood concepts and representations, because they provide the strongest kind of evidence for my case. Whatever view you take of our conceptual repertoire, given our meta-representational ability, there must be concepts and representations our grasp of which is such that we can *think about them* without being able to *think with them*. In other terms, there must be concepts that we can meta-represent, without being able to deploy them to represent the object or properties they are the concepts of. Similarly, there must be representations that we can meta-represent, without being able to deploy them to represent the states of affairs that would make them true. Such concepts and representations may occur in a meta-representation embedded in a validating context, yielding an undisquotable reflective belief. Ergo reflective beliefs are a stable ingredient of our mental life.

Once you have had to accept the existence of reflective beliefs in order to handle the problem raised by partially understood beliefs, you might as well see what further ligh this might throw on human cognition. I want to suggest that reflective beliefs do, in fact, play a major role in human cognition, and that not just partially understood beliefs, but also many well understood beliefs are reflective beliefs, paradigmatic examples being scientific beliefs. Some of the concepts that are used in scientific claims are well-understood by scientists, but they may remain beyond the reach of their intuitions. These are concepts that scientists can indeed *think with*, but, in most cases, only *by thinking about them*, that is, only reflectively. Typically, the validating contexts of beliefs containing such scientific concepts are not (for competent scientists) in the form of a reference to an authority, but in the form of an argument or a demonstration. Such arguments and demonstrations are not of a kind delivered by spontaneous inference, and must be reflected upon in order to see their force.

Of some concepts, we have an intuitive grasp. We can use them unreflectively, and normally without running into paradoxes (unless a philosopher is intent on pushing us where we would not spontaneously go). Of other concepts, we may have only a limited grasp. Of yet other concepts, we may have a full grasp, but a grasp of a kind that can only be deployed reflectively. Not every kind of concept can enter into every kind of belief.

Intuitive and reflective concepts

Many of our beliefs are grounded in perception and in spontaneous and unconscious inference from perception. You see autumn leaves under a bare tree, and you spontaneously form the beliefs:

(30) There are leaves under this tree.
(31) This tree has lost its leaves

You hear the doorbell ringing, and you spontaneously form the beliefs:

(32) The doorbell is ringing.
(33) There is someone at the door.

You see your friend Martha frowning, and you spontaneously form the beliefs:

(34) Martha is frowning
(35) Martha is worried about something

When a child, you were shown a bird and told: “this is a sparrow”. You spontaneously formed the beliefs:

(36) This is a bird
(37) Sparrows are birds

In all four cases, the first belief is based on perception, the second on inference. If challenged, you might be able to produce, ex post facto, a missing premise which, together with the perceptual belief, warrants the inferential belief. However, the fact is that you arrived at the inferential belief without engaging in deliberate or conscious inference. All beliefs that are the output of perceptual processes are intuitive in a standard psychological sense, and so are all beliefs that are the output of spontaneous and unconscious inferential processes taking intuitive beliefs as premises. Intuitive beliefs can be activated in our mental life without activating or even remembering the premises from which they were derived. You probably do believe, for instance, that sparrows are birds, but do you remember whether you formed this belief by generalisin from an instance shown to you, or by being told about sparrows?

What is the conceptual repertoire on which intuitive beliefs draw? Perceptual devices take sensory data as input and deliver as output conceptual identifications of the distal stimuli that caused the sensation. To do this, perceptual devices must draw on a conceptual repertoire which contains all the concepts of the things that can be perceptually identified. Spontaneous inferential processes derive intuitive beliefs from perceptual beliefs and from other inferentially derived intuitive beliefs. Some inferentially intuitive beliefs are about things that cannot be directly perceived. These inferential processes must therefore be able to draw on a conceptual repertoire that goes beyond that of the perceptual processes. If we take together the conceptual repertoires of perceptual processes and of spontaneous inferential processes, we get the conceptual repertoire of intuitive beliefs, or, in other terms, a repertoire of intuitive concepts.

Rather than arriving at intuitive beliefs such as (30)-(37) by means of your own perceptions and inferences, you might have arrived at them via communication. Someone you trust might tell you any of (30)-(37). You would then disquote the content of the communication from the belief that it has been communicated and believe this content directly. Communication plays here, to some extent, the role of experience by proxy. You might yourself have formed such beliefs via perception and spontaneous inference, had you been placed in a position to experience their perceptual basis. Whether acquired directly or by proxy, intuitive beliefs must be exclusively represented by means of the intuitive repertoire.

Not all our beliefs are intuitive in the sense of being derived from perception plus spontaneous inference, either directly or by proxy. Consider (38)-(42):

(38) The prime factors of 9139 are 13, 19, and 37.
(39) Water is H_2O.
(40) A king and two knights are not enough to force a checkmate.
(41) Where the judiciary is not independent, there can be no true democracy.

(38)-(41) are examples of propositions that may be understood and believed. In ordinary circumstances and for most people at least – I am hedging in order to avoid orthogonal issues to which I will soon revert -, the states of affairs that make these propositions true cannot be perceived, but only inferred, and, moreover, inferring them requires some conscious and deliberate thinking. Often, such beliefs are acquired not via ratiocination, but via communication. But communication in this case does not provide one with experience by proxy. No experience could have triggered the spontaneous formation of these beliefs anyhow. These beliefs, then, are not, at least for most of us, intuitive beliefs in the intended sense. Moreover, in most cases, when such non-intuitive beliefs serve as premises in the derivation of further beliefs, the inferential processes in which they are involved are conscious and deliberate ones. Non-intuitive beliefs typically beget further non-intuitive beliefs.

The conceptual repertoire of non-intuitive beliefs is richer than the intuitive repertoire. You may learn a new concept, say in mathematics, chemistry, philosophy, or chess, understand it properly, be able to make it work for you in constructing hypotheses or arguments, but not be able to draw spontaneous inferences on the basis of its occurrence in a conceptual representation. The inferences in which such concepts are involved are performed by invoking explicit schemas or procedures. Some mathematical geniuses may “see” prime numbers as primes, and infer intuitively the prime factors of many numbers. Most of us may know quite well what a prime factor is but be able to discover the prime factors of 9139 only by painstaking calculations. A chess expert may “see” all checkmate positions, intuit all the possible outcomes of endgames, and therefore believe (40) intuitively. A chess beginner, on the other hand, may have properly understood the concept of a checkmate, but have to verify every instance; for such a beginner, concluding that a king and two knights are not enough to force a checkmate involves testing, one by one, alternative moves.

Of some concepts, we have a reflective mastery, but no intuitive grasp. We understand them because we have beliefs about them. These “reflective concepts”, as I will ca them, are introduced by explicit theories which specify their meaning and the inferences that can be drawn on their basis. The possibility of such reflective concepts follows from the existence of the human meta-representational ability.

Of course, a concept may start its life in the mind of an individual as a reflective concept and later become an intuitive one, that is, come to determine spontaneous rather than deliberate and conscious inferences. For instance, when you were first taught, as a child, the difference between odd and even numbers, “odd” and “even” were probably reflective concepts for you: you had to think hard in order to decide whether a given number was odd or even, or what followed from a number being such. By now, I assume that these two concepts have become quite intuitive to you. Incidentally, the reverse movement from intuitive to reflective is also possible: for instance, you may have had an intuitive grasp of the concept of weight, which was at least temporarily lost when you became aware of the necessity to distinguish weight and mass. You may then have deployed the concept of weight reflectively, at least in classroom contexts. Such movement from intuitive to reflective status and back may be involved in concept revision (in which case, it is not a simple matter to decide whether, strictly speaking, the *same* concept is involved throughout).

How easily, in the cognitive life of the individual, may reflective concepts become intuitive concepts? We may characterise the range of possible answers by sketching two extreme ones, a radically empiricist one and a radically nativist one. According to a radically empiricist answer, all concepts are acquired by learning the words that encode them. No concept is immediately intuitive, but all concepts may become intuitive, provided they are used often enough (just like tying shoe laces is not immediately intuitive but becomes so with practice). According to a radically nativist answer, there is an innate range of intuitive concepts, a subset of which becomes actualised in the intuitive repertoire of any given individual. An individual’s intuitive repertoire does not contain all the concepts in this innate range, but it cannot contain concepts outside that range Therefore, reflective concepts that fall outside the innate range can never become intuitive concepts. I assume that these two extreme views are wrong and that truth lies somewhere in between. Susan Carey (1985, 1991), for instance, has developed a view more on the empiricist side. I have explored a view more on the nativist side (Sperber 1994b, 1996). All this remains, for the time being, quite speculative.

The question of the relative fluidity, viscosity, or rigidity of the intuitive and reflective repertoires is a fascinating one, but it is orthogonal to the argument of the present paper. Even if the extreme empiricist view were correct, and all reflective concepts were capable of becoming intuitive ones, it would still be the case that, at any given time in the cognitive life of an individual, some concepts are mastered only in a reflective way, and therefore the beliefs into which these concepts enter cannot be intuitive beliefs. Assuming that there are only two categories of beliefs, intuitive and reflective, then intuitive concepts can occur both in intuitive and reflective beliefs, and reflective concepts can occur only in reflective beliefs.

Believing the same thing twice

The difference between intuitive and reflective beliefs is one of mental inscription, and not necessarily of content. A representation involving only intuitive concepts may be believed intuitively or reflectively. It should even be possible that some contents be believed both intuitively and reflectively by the same individual at the same time, each belief playing a different role in the believer’s thinking and behaviour.

One may come to hold a belief both intuitively and reflectively by adding to the data-base a copy of a reflective belief disembedded from its validating context. Consider again the railway example (21)-(22). You went to the railway employee and, on the basis of what he told you, you formed two intuitive beliefs, the second being disembedded from the first:

(21) The railway employee (who is to be trusted on such matters) said that the last Sunday train to Oxford is at 11:45.

(22) The last Sunday train to Oxford is at 11:45.

If you kept the meta-representational belief (21) in memory, you thereby kept a reflective belief with content (22), that is, you were in a position to assert (22) or use it in your thinking not only as a plain data-base belief but also as a reflective belief, at least implicitly connected to its validating context. Suppose that after coming back from the information desk, you had the following dialogue with your travelling companion:

(42) *She*: So, when is the last Sunday train to Oxford?
You: It is at 11:45.

In answering, you were doing two things. You were reporting what the employee told you, and you were expressing your belief that the last Sunday train to Oxford is at 11:45. You were expressing your belief *by* reporting the employee's utterance: your echoing his utterance without any reservation amounted to endorsing what you were echoing. In this case, your utterance is best seen as an expression not of your intuitive belief, but of your reflective belief: you were speaking under the authority of your informant. Months later, you might well remember that the last Sunday train to Oxford is at 11:45, but not anymore that you were told so by a railway employee.

One may also take a reflective stance to a belief that, initially, was purely intuitive, and come to believe it reflectively too. The belief that the Earth is flat has been, no doub a widespread intuitive belief in human history. In fact, the intuitive pull of this belief is such that, for most people who know it to be false, it still takes a moment reflection to understand that an aeroplane could not cross the Atlantic in a straight line. In the sixth century, the monk Cosmas, author of a *Christian Topography*, believed that the Eart was flat (see Wilford 1981, chapter 4). I assume that he believed it intuitively. Before becoming a monk, he had been a great traveller and must have drawn on this intuitive belief when thinking about his past and future journeys. It is clear from his writings that Cosmas came to believe in the flatness of the Earth, not just intuitively, but also reflectively. He provided several validating contexts for this belief, some based on the authority of the Scriptures, others based on rational arguments, such as the absurdit of the very idea of antipodes. Cosmas used his reflective belief that the Earth was flat to construct his *Topography*, and to argue against those who, following the Greeks, believed that the Earth was spherical.

Reflective beliefs being a broad and loose category of subtly different credal attitudes, the same content may come to be reflectively believed in different ways. A student may believe a mathematical theorem in a validating context consisting in an acknowledgment of the authority of her teacher. She may later understand the proof and use it as a validating context for the same content.

Mathematicians themselves may come to believe the same proposition in a different but still reflective way. Consider, for instance, Fermat's conjecture, which was firmly believed to be true – more firmly than most empirical beliefs – before it was actually proven in 1994. Still, the belief has changed, not in its content, but in its validating context. It used to be something like (40), and now it is something like (41):

(40) It is a very-well-supported conjecture that, for $n>2$, $x^n+y^n = z^n$ cannot be solved in integers x, y, z, with $xyz \neq 0$.

(41) It is a theorem that, for $n>2$, $x^n+y^n = z^n$ cannot be solved in integers x, y, z, with $xyz \neq 0$.

This change in validating context goes together with a profound change in the epistemic role that this reflective belief may play in mathematical reasoning. Fermat's conjecture was something to be proven; it can now be used as a premise in deriving new theorems (and, for all consequences of Fermat's conjecture that were already known, the proof changes them into absolute theorems).

***

To wrap it up:

We have two kinds of beliefs. We have intuitive data-base beliefs, which are inscribed in our mind in a manner such that they are automatically treated as data. They are expressed in an intuitive mental lexicon that allows spontaneous inference. Intuitive beliefs are a most fundamental category of cognition. Given the fact that we have intuitive beliefs and a meta-representational ability, we are also capable of having reflective beliefs and reflective concepts, or to take a reflective stance towards intuitive concepts and beliefs. Reflective beliefs are a loose family of derived attitudes that are continuous with other reflective attitudes of a non-credal kind.

While reflective beliefs, unlike intuitive beliefs, are not a basic category of cognitive architecture, they play a major role in the development and transmission of cultural representations, allowing concepts and ideas that are only half-understood, or that are well understood but only within the context of explicit theories, to stabilise in a human population and to expand the range of thoughts that can be entertained, way beyond what would be possible on a strict intuitive basis. It is arguable (see Sperber 1996) that much of culture, from religion to science, is made of reflective concepts and beliefs.

References

Carey, Susan (1985). *Conceptual Change in Childhood*. Cambridge, Mass: MIT Press.

Carey, Susan (1991). Knowledge Acquisition: Enrichment or Conceptual Change? In S. Carey and R. Gelman (eds.) *Epigenesis of Mind: Studies in Biology and Cognition*. Hillsdale: Erlbaum.

Needham, Rodney (1972). *Belief, Language and Experience*. Oxford: Blackwell.

Recanati, François (1997) Can We Believe What We Do Not Understand? *Mind and Language* ….

Sperber, Dan (1975). *Rethinking Symbolism*. Cambridge: Cambridge University Press.

Sperber, Dan (1982/1985) Apparently Irrational Beliefs, In S. Lukes & M. Hollis (eds.), *Rationality and Relativism* (Oxford, Blackwell, 1982). Revised version in D. Sperber (1985). *On Anthropological Knowledge*. Cambridge: Cambridge University Press.

Sperber, Dan (1990). The Epidemiology of Beliefs. In Colin Fraser & George Gaskell (eds.) *The Social Psychological Study of Widespread Beliefs*. Oxford: Clarendon Press.

Sperber, Dan (1994a). Understanding Verbal Understanding. In Jean Khalfa (ed.) *What is Intelligence?* Cambridge University Press.

Sperber, Dan (1994b). The Modularity of Thought and the Epidemiology of Representations. In L. A. Hirschfeld & S. A. Gelman (eds), *Mapping the Mind: Domain Specificity in Cognition and Culture,* New York: Cambridge University Press, 39-67.

Sperber, Dan (1996). *Explaining Culture: A Naturalistic Approach*. Oxford: Blackwell.

Sperber, Dan & Deirdre Wilson (1981). Irony and the Use-Mention Distinction. In P. Cole (ed.) *Radical Pragmatics* (New-York, Academic Press, 1981) 295-318.

Sperber, Dan & Deirdre Wilson (1986/1995). *Relevance: Communication and cognition*. Oxford: Blackwell, 1986. *Second Edition,* Oxford: Blackwell 1995.

Wilford, John N. (1981). *The Mapmakers*. New York: Alfred Knopf.

Wilson, Deirdre & Dan Sperber (1992) On Verbal Irony. *Lingua* 87, 53-76.