

Expressivism about Belief

Philipp Keller*

June 18, 2012

The basic puzzle: a sentence such like “the weather is fine” may be used not only to talk about the weather, but also to talk about us – it may successfully be deployed to characterise Sam as believing, desiring, wishing, Maria as hoping, expecting or remembering and Fred as saying, stating, judging, supposing or asking. This use of sentences about the weather in the classification and characterisation of mental states raises a problem: what is the relation between these states and the state of the weather called upon to characterise them? In virtue of what are these characterisations true?

Many people have thought that a useful first step towards answering these questions is to focus on just one type of mental state, belief, and one type of characterisation, the stating of truth-conditions. This approach has a number of drawbacks:

- Starting with belief is insufficiently general: beliefs are just one kind of mental states, and the conceptual apparatus that is useful to analyse them may be too parochial to be of any help in the analysis of other mental states.
- Taking beliefs as paradigm mental states may tempt us to reify truth-conditions into self-standing objects, propositions or contents. These propositions, introduced as objects of beliefs, are too fine-grained to be plausibly taken to be the objects of the other ‘propositional’ attitudes.
- Neglecting other mental states than beliefs, we misconstrue the phenomenal, normative and affective properties of the latter: they come to be seen as phenomenally, normatively and affectively inert, and we end up with a meagre and lifeless conception of our desire for, and love of, truth.
- Focussing too much on belief, we make it hard to understand how belief connects to other mental states. Especially mental states that are entanglements of belief with other states fall out of the picture.

Rather than starting with beliefs, we should first develop a general theory of correctness conditions. The fundamental notion, I submit, is representation-as. The content-conditions for representation-as have a two-fold structure: To represent something *a* as *F*, I must be *appropriately related* to a external item (a proposition, a state of affairs, a perceptual situation, an object, a plan) and my standing in this relation to the external item must be an appropriate *reaction* to it (serious/assertoric/affirmative, non-lucky, veridical, non-perverse, not practically irrational). The appropriateness of the reaction is thus the joint outcome of two factors:

- it must be caused by (and justifiable in terms of) an appropriate *cognitive base*, which is directed at the representational / direct objects and represents them as being a certain way;
- it must be accompanied by an *affective feeling* that is the right kind of reaction to the right kind of intentional / formal object.

The double structure of representational states, reaching out to something beyond themselves and at the same time situating us with respect to it, allows for two types of failure to meet the characteristic

*Département de Philosophie, Université de Genève, Switzerland, philipp.keller@unige.ch.

content-conditions for the respective state:

1. Judgements have success-conditions: My judgement of a to be F is successful iff I succeed in referring to a and to predicate of it to be F . Two types of failure are possible:
 - (a) Wrong direct object: I do not refer to a (but to something else or to nothing at all) or do not predicate of it to be F (but to be something else, or nothing at all).
 - (b) Wrong formal object (misfiring of the speech-act): I do not really make a judgement, but perform some other speech-act or no speech-act at all.
2. Beliefs and states of knowledge have truth-conditions: My belief that a is F is true iff Fa . Beliefs are criticisable on two grounds:
 - (a) Wrong direct object: it is not true that a is F .
 - (b) Wrong formal object: my belief falls short of knowledge, i.e. has not the right genesis to be a manifestation of my desire for truth.
3. Perceptions have veridicality-conditions: My perception of a as F is veridical iff I am visually presented to a and see it as F . Two types of perceptual ‘error’ may be distinguished:
 - (a) Wrong direct object (illusion): While I do have a perceptual experiences as of a being F , I am either not visually presented to a or do not see it as F .
 - (b) Wrong formal object (hallucination): My state cannot be characterised as one of perceiving.
4. Emotions have appropriateness-conditions: My emotion of type T towards a is appropriate iff a exhibits the formal object of emotions of type T .
Two types of error:
 - (a) Wrong direct object: My fear is not directed at a dog that is in fact dangerous.
 - (b) Wrong formal object: I see the dangerous dog, but I don’t see it as dangerous. While I am struck by the right value (dangerousness), I don’t react in the right way (by fear).
5. Desires have satisfaction-conditions: My desire to ϕ is satisfied iff I want to ϕ and I ϕ .
Two types of error:
 - (a) Wrong direct object: I don’t ϕ .
 - (b) Wrong formal object: I ϕ , but I don’t do it as a my plan. While I realise my plan, I don’t react in the right way to my ϕ -ing, i.e. I don’t see it as the satisfaction of my desire.

The arguments in favour of objects of belief

Frege, Bolzano and the contemporary orthodoxy hold that belief is a propositional attitude, that to believe that p is to stand in a relation to the referent of “that p ”.¹

The most important argument in favour of a relational construal of belief is that we need it to explain the similarity between John’s believing that Sam is Maria’s wife and Sam’s believing that snow is white. If this are just two states of John and Sam respectively, the argument goes, how are we to explain that they make both of them believers? This argument is no good: by the same token, you could say that the similarity between my running at ten kilometres per hour and your running at fifteen kilometres per hour is an argument for construing running as a relation between a runner and a speed. The same example rebuts another argument in favour of a relational account of belief: while it is true that if someone is a believer, then s/he believes something, the soundness of existential generalisation is not conclusive evidence of relationality. It is equally necessary that if you run, you run at a certain speed – but, again, this does nothing to show that running is a relation.

More explicitly, it has been taken to be an argument in favour of the acceptance of objects of belief that we can infer both (4) and (5) from (1-3):

¹Bach (1997: 222–223) cites Burge (1980: 55), Fodor (1978: 178) (cited after reprint in Fodor (1981)), Schiffer (1992: 491,505), Soames (1987: 105–106) and Stalnaker (1988: 140–141) as representatives of the orthodox view.

- (1) Huey believes that snow is white.
- (2) Dewey believes that grass is green.
- (3) Louie believes that grass is green.
- (4) Both Huey and Dewey believe something.
- (5) There is something that both Dewey and Louie believe.

More generally, we can infer (7) from (6) – allegedly showing that believing is a relation to some object of belief:

- (6) Huey is a believer.
- (7) There is something that Huey believes.

Both these arguments are inconclusive, however. For we can also, and with equal right, both infer (4*) and (5*) from (1*-3*) and infer (7*) from (6*):

- (1*) Huey runs at 8 km/h.
- (2*) Dewey runs at 10 km/h.
- (3*) Louie runs at 10 km/h.
- (4*) Both Huey and Dewey run at a certain speed.
- (5*) There is some speed at which both Dewey and Louie run.
- (6*) Huey is a runner.
- (7*) There is some speed at which Huey runs.

But running is not standing in a relation to a speed. What makes the inferences valid, is a certain metaphysical necessity: that every running occurs at a certain speed means that running always allows for adverbial modification with a determinate of the determinable *SPEED*. This does not, however, mean that running has an internal object which is a speed.

In the same way, it seems to me, it may very well be true that if someone believes, there is always something that is believed, even though believing is not a relation to what is believed.

Belief and acceptance

Apart from questioning its motivations, I think we can give a positive argument against the relational construal of belief. The argument proceeds from an analogy with truth and truthmaking. In the same way that the truth of truth-bearers (here understood to be interpreted sentences) is a de-relativisation of their being made true by something, belief is a de-relativisation of a conceptually prior relation of acceptance. It is true that in order to believe, you have to accept something (an interpreted sentence, I will argue), in the same way than it is true that in order for something to be true, it has to be made true by something. But there are good reasons to think that this does not turn truth into a relation – by analogy, there are good reasons to think that to believe is not to stand in a relation to anything.

If believing that *p* is a disposition to credally feel that *p*, accepting that *p* is an individual mental act or as a more general policy – it is something people do, rather than just a state they are in. As a mental act, it is subject to act-specific norms that may depend on the reasons some subject has for his act.²

Truthmaker maximalists hold that truth is a derelativisation of the relation of being made true, though not itself relational. The things made true are plausibly taken to be sentence-tokens or inscriptions, that are made true in some specific way: The inscription “Huey ate the cookies on the couch”, e.g., is made true by some cookie-eating event as an inscription of English and as specifying the location of either Huey or the cookies.

²Minimally, an acceptance of *p* involves the commitment to use *p* as a premise in reasoning and decision-making. You can, however, wilfully and deliberately accept as a premise a proposition that you do not believe to be true, i.e. on a prudential rather than on an evidential basis.

Analogously, an adverbial theory of belief might hold that belief is a derelativisation of a broader and conceptually prior notion of acceptance, in which sentence-tokens stand to believers in certain ways. Believing that p is a matter of there being a sentence-token that is accepted in a certain way (i.e. the way in which, if it were true, would be made true).

Two immediate problems arise:

1. The famous translation argument, going back to Frege and expanded by Church (1950) and Bigelow (1980) against sententialist and paratactic theories of belief ascription seems applicable: (i) if the relations of acceptance are sentence-tokens, (1) and (8) Tick glaubt, dass Schnee weiss ist. involve relations to different sentences. Moreover, (ii) the acceptance of “Snow is white” by Huey can only underlie the truth of (1) if we presuppose that, in it, “snow” means snow rather than, say, grass. But the claim is not that (9) Huey accepts some sentence-inscription which means that snow is white. is synonymous with (1), or conceptually equivalent with it, or an explication of what is implicit in (1). (9) rather specifies some truthmaker for (1): it is a metaphysical analysis of (1) and their correlation a substantial thesis.
2. The adverbial theory seems committed to (i) ascribing beliefs only to language-using creatures and (ii) only in contents for which some inscription may be found. Against (ii), the adverbialist may claim that the accepted inscription is produced in (1). To avoid (i), the relation of acceptance has to be construed as passive (Cohen 1992) and as a more general attitude than belief as ordinarily construed, which “comes apart from belief in cases where one is warranted in acting on the assumption that p or taking it for granted that p or trusting that p for reasons that do not bear on the likely truth of p ” (Wright 2004: 177). So construed, we may say that (intelligent) dogs accept sentence-tokens and accordingly have beliefs.

A step in the right direction: Gricean intentions

A theory of belief as a derelativisation of a prior relation of acceptance of interpreted sentence-tokens leaves many questions open. In particular, it does not explain how our beliefs acquire content, and acquire the content they do.

In his seminar work on non-natural meaning, Grice (1957 1967 1968 1969 1978 1982) attempted to provide an answer to these questions: It is because some utterance is produced with a certain intention that it has meaning (Grice calls it “occasion-meaning”) and that it has the meaning it does. If the Gricean programme works, we explain representationality in terms of intentionality: Utterances acquire their representational properties in virtue of standing in some causal connection to mental states that have intentional properties.

How the speakers’ intentions acquire *their* content, however, is a question left unaddressed.

A step in the right direction: Use-theoretic accounts of meaning

Use-theories of meaning, as recently advocated e.g. by Paul Horwich, aim to address this question, showing how the ascription of certain meaning intentions best rationalises our usage of some words. Their problem, however, is the converse one: they succeed in giving an reductionist account of meaning only to the extent that we may identify the (pattern of) uses of some given word without knowing its meaning, and even without presupposing that it has meaning at all. This assumption, I argue, can

plausibly be contested.

According to the minimalist theory of truth, grasping the concept of truth is to have the disposition to underivedly accept all (or at least, all but the Liar-like) instances of the so-called “equivalence schema”:

(1) the proposition that p is true $\leftrightarrow p$

Underived acceptance is acceptance that does not stem from the acceptance of other sentences containing the truth-predicate. Because we do not want to say, as Tarski has to, that our concept of truth changes with every change to our language, instances of (1) have to be accepted *in virtue of their having a certain form*, i.e. in virtue of their being instances of the schema (1). This means that everyone having the concept of truth must have the concept of material equivalence \leftrightarrow . This implausible consequence of the minimalist theory could be avoided if we required that someone possessing the concept of truth has a disposition to accept all (unproblematic) instances of the following inferences:

(2)
$$\frac{\text{the proposition that } p \text{ is true}}{p} \qquad \frac{p}{\text{the proposition that } p \text{ is true}}$$

The problem with (2), however, is that underived acceptance of an inference is less clear a notion than underived acceptance of a sentence: while the primitively compelling introduction rules for logical constants add deductive power, nothing of that kind seems to happen with the trivial reformulation in (2).

The minimalist claim about the meaning of “true” is but a special case of a more general heuristic to determine the property constituting the meaning of some given expression type w given by the “use theory of meaning”:

1. Identify the overall pattern of use of w within a linguistic community C .
2. Identify the core (= meaning-constituting) use of w (the part of the overall-use of w that is explanatorily basic).
3. Find an idealised law that best systematises the core use of w within C : specify the sentences (conditional) acceptance of which is, within C , constitutive of mastery of the concept expressed by w .

The basic acceptance property of a word, the “common factor in the explanations of its numerous occurrences” (Horwich 2005: 26) is then identified with the meaning with the meaning of the word.

While use-theoretic accounts of meaning certainly constitute progress with respect to the (basic form of) the Gricean programme, they presuppose that we can identify words without first knowing their meaning: that there is a pattern of use of some individuable thing, that we can abstract from to say what kind of thing it is.

The promise of a general solution? Expressivism about Belief

Here, I think, is where expressivism about belief can come to the rescue. Suppose we agree that normative sentences express attitudes. We then have to answer the question why some given sentence expresses the attitude with content x rather than an attitude with content y . The same-content theorists have an answer: because the sentence has content x (rather than y). The expressivists cannot say this, because in their view the sentence has content x *because* the attitude has the content x :

...expressivists are committed to holding that the sentence 'grass is green and snow is white' gets its content *from* the belief it expresses. But in order for that to happen, it must *first* express a belief, and only *then* acquire a content. (Schroeder 2008: 25)

Expressivists hold that evaluative sentences express, but do not report attitudes. But how do we know, of a given evaluative sentence, which attitude it expresses? To know this, we have to know the content of the attitude, as attitudes are individuated (at least partly) by their content. Two problems for expressivists:

- epistemic determination: in order to know the content of the attitude, we need to know the content of the sentence; but we cannot know this (e.g. by truth-conditions), *before* we know the content of the attitude;
- metaphysical determination: the sentence has the content it has because the attitude has the content *it* has, the sentence inherits its content from the attitude; hence, the content of the attitude cannot itself be determined by the content of the sentence.

How are these problems to be solved? Assertability semantics.

The sentence gets its content autonomously, in virtue of being associated with a proposition that the speaker is in a certain mental state. Which proposition this is (i.e. which belief the speaker required to have) is determined by its assertability conditions. These conditions require the speaker to have the attitude in question, but do not state *that* the speaker has this attitude. These assertability conditions are different from the truth-conditions:

On the picture to which expressivists are committed [...] [t]he primary job of the semantics is to assign to each atomic sentence a mental state – the state that you have to be in, in order for it to be permissible for you to assert that sentence. (Schroeder 2008: 33)

Back to the negation problem

To solve the negation problem, expressivists have to assume that “intention is subject to a basic, noncognitive kind of inconsistency, the kind of inconsistency on which expressivists ought to be able to model their account of the inconsistency between the mental state expressed by ‘murder is wrong’ and that expressed by ‘murder is not wrong’” (Schroeder 2008: 43).

The negation problem: To explain why the attitude of ϕ ing murdering ascribed to Jon in (n2') is inconsistent with the one ascribed to him in (w'):

- w' Jon disapproves of murdering.
- n1' Jon does not disapprove of murdering.
- n2' Jon ϕ s murdering.
- n3' Jon disapproves of not murdering.

The need to provide for a ϕ for (n2') is motivated by the suggested analogy with:

- w Jon thinks that murdering is wrong.
- n1 Jon does not think that murdering is wrong.
- n2 Jon thinks that murdering is not wrong.
- n3 Jon thinks that not murdering is wrong.

Problem is that we need a sentential, not just a predicational operator, allowing for a three-fold distinction between $\neg\Box p$, $\Box\neg p$ and $\neg\Box\neg p$. Hence, we need dual operators. But such dual operators are readily available: The distinction between (n2) and (n1) is the one between:

- disapproval / tolerance
- not recommending / advising against
- agnosticism / atheism
- non-acceptance / rejection
- obligation / permission
- hope / regret
- anticipation / memory

For example:

- v' Jon recommends giving.
- m1' Jon does not recommend giving.
- m2' Jon advises against giving.
- m3' Jon recommends not giving.

compared to

- v Jon thinks that it is rational to give.
- m1 Jon does not think that it is rational to give.
- m2 Jon thinks that it is not rational to give.
- m3 Jon thinks that it is rational not to give.

There are four types of inconsistency:

1. logical: the opposition between (w) and (n1) and between (w') and (n1'), due to the semantics of "¬".
2. objectual: the opposition between (w) and (n3) and between (w') and (n3'), due to the fact that the same attitude takes 'contrary' objects; why these objects are contrary has further, case-specific reasons (for actions: that they cannot be done both by the same person at the same time; for plans: that they are not realisable at the same time)
3. "type A": analogous to the cases of belief and intention ('personal inconsistency')
4. "type B": neither of the other three

Why expressivists need not be too much troubled by type B 'inconsistency':

- Whether or not an opposition is of type B, depends on what we count as analogous to the opposition between " Bp " and " $B\neg p$ ".
- We may deny that (n2') is in the relevant sense an evaluative sentence. Ethics prescribes, deals with what is obligatory and what is forbidden; it only deals indirectly with what is permissible, by staying silent about it.
- The predicate that needs analysis in (w') is 'thinking-to-be-wrong': this is an evaluative predicate, because it motivates, expresses a desire-like state etc. The 'contrary' predicate, 'thinking-to-be-not-wrong', however, is not motivating.
- Compare the case with imperatives, desires and promises:
 - v'' Close the door! // I desire to p . // I promise to ϕ .
 - m1'' I don't order you to close the door. // I don't desire to p . // I don't promise to ϕ .
 - m2'' ?? // ?? // ??
 - m3'' Don't close the door. // I desire not to p . // I promise not to ϕ .

References

- Kent Bach, 1997, "Do Belief Reports Report Beliefs?", *Pacific Philosophical Quarterly* 78, pp. 215–241.
- Ned Block (ed.) 1981, *Readings in the Philosophy of Psychology*, volume II, Cambridge, Massachusetts: The MIT Press.
- Tyler Burge, 1980, "The Content of Propositional Attitudes", *Noûs* 14(1), pp. 53–58.
- Peter Cole and Jerry L. Morgan (eds.) 1975, *Syntax and Semantics 3: Speech Acts*, New York: Academic Press.
- Donald Davidson and Gilbert H. Harman (eds.) 1975, *The Logic of Grammar*, Encino and Belmont, California: Dickenson Publishing Co.
- Jerry A. Fodor, 1978, "Propositional Attitudes", *The Monist* 61(4), pp. 501–523, reprinted in Block (1981) and Fodor (1981).
- Jerry A. Fodor, 1981, *Representations: Philosophical Essays on the Foundations of Cognitive Science*, Cambridge, Massachusetts: The MIT Press.
- H. Paul Grice, 1957, "Meaning", *The Philosophical Review* 66, pp. 377–388, reprinted in Grice (1989).
- H. Paul Grice, 1967, "Logic and Conversation", The William James Lectures, delivered at Harvard University in 1967, and thereafter distributed in unpublished form for many years. Portions published in Davidson and Harman (1975: 64–75) and Cole and Morgan (1975: 41–58). The whole published, w.
- H. Paul Grice, 1968, "Utterer's Meaning, Sentence-Meaning, and Word-Meaning", *Foundations of Language* 4, pp. 225–242, reprinted in Grice (1989).
- H. Paul Grice, 1969, "Utterer's Meaning and Intentions", *The Philosophical Review* 87, pp. 147–177, reprinted in Grice (1989).
- H. Paul Grice, 1978, "Further Notes on Logic and Conversation", in Peter Cole (ed.) "Syntax and Semantics 9: Pragmatics", pp. 113–127, New York: Academic Press, reprinted in Grice (1989).
- H. Paul Grice, 1982, "Meaning Revisited", in N.V. Smith (ed.) "Mutual Knowledge", pp. 223–243, London: Academic Press, reprinted in Grice (1989).
- H. Paul Grice, 1989, *Studies in the Way of Words*, Cambridge, Massachusetts: Harvard University Press.
- Paul Horwich, 2005, *Reflections on Meaning*, Oxford: Clarendon Press.
- Stephen Schiffer, 1992, "Belief Ascription", *The Journal of Philosophy* 89, pp. 499–521.
- Mark Schroeder, 2008, *Being For: Evaluating the Semantic Program of Expressivism*, Oxford: Clarendon Press.
- Scott Soames, 1987, "Substitutivity", in Judith Jarvis Thomson (ed.) "On Being and Saying: Essays For Richard Cartwright", pp. 99–132, Cambridge, Massachusetts: The MIT Press.
- Robert C. Stalnaker, 1988, "Belief Attribution and Context", in Robert Grimm and Daniel D. Merrill (eds.) "Contents of Thought", pp. 150–166, Tucson, Arizona: University of Arizona Press, reprinted in Stalnaker (1999: 150–166).
- Robert C. Stalnaker, 1999, *Context and Content: Essays on Intensionality Speech and Thought*, Oxford: Oxford University Press.