

# Private Investigators and Public Speakers

Alexander Sandgren

*forthcoming in The Journal of the American Philosophical Association*

preprint, please cite published version

## Abstract

Near the end of 'Naming the colours', Lewis (1997) makes an interesting claim about the relationship between linguistic and mental content; we are typically unable to read the content of a belief off the content of a sentence used to express that belief or vice versa. I call this view *autonomism*. I motivate and defend autonomism and discuss its importance in the philosophy of mind and language. In a nutshell, I argue that the different theoretical roles that mental and linguistic content play suggest these kinds of content should be understood as sensitive to different things.

## 1 Horses for Courses

Near the end of 'Naming the colours', Lewis (1997) makes a provocative proposal about the relationship between the content of language (e.g. what is said by speakers making utterances) and mental content (e.g. the content of beliefs and perceptual representations). Lewis suggests that linguistic and mental content are, in a crucial sense, autonomous (though he does not use that word). If this suggestion is right, one will typically not be able to read the content of a belief off the content of a sentence used to express that belief or vice versa, *even in the simplest cases*. For example, when I express my belief by saying 'the Earth goes around the Sun', the content of the belief I express and the content of my assertion are not guaranteed to be the same. Indeed they are likely not to be the same since the content of my utterance and the content of my belief are determined by different things and these different determiners tend to pull in opposite directions. Let's call this view *autonomism*. Lewis gives only a minimal direct defence of autonomism, though some of the remarks about belief content and belief reports in section 1.4. of 'On the Plurality of Worlds' can plausibly be read as a partial motivation for autonomism (see especially Lewis (1986: 27-36)).

Lewis's suggestion and autonomism have not received the attention they deserve. The purpose of this paper is to clarify, motivate, and defend autonomism and to discuss some of its import in the philosophy of mind and language. The motivation for autonomism I will focus on is based on the different theoretical roles that linguistic content and mental content are supposed to play. In a nutshell, these different roles suggest that mental content and linguistic content should be understood as *sensitive* to different things since the sensitivities that make one kind of content fit for its role make the other kind of content ill-fitting its role, and vice versa.

## 2 Content

I will work with fairly rough conceptions of linguistic content and mental content, setting a number of issues aside. Linguistic content will be understood as what is said by the speaker when they make the utterance. So if I say ‘the meeting will start at 11’, the *content* of my utterance is what I said by making that utterance. The distinction between semantics and pragmatics can be harmlessly set aside for present purposes. I will also set aside the distinction between the content of whole utterances and the content of individual expressions. Nothing in the present discussion turns on these distinctions or where they are drawn.

The content of a belief, we will say, is just how the believer represents things to be in virtue of having that belief. So if I believe that the meeting will start at 11, then I represent the world as being one in which the meeting starts at 11.<sup>1</sup> When discussing mental content I will focus on belief, but we can think of the content of other attitudes in a similar way.

### 2.1 Linguistic Content

Next, let’s consider a broad question: what is language *for*? Most obviously, language allows us to communicate, ask and answer questions, express disagreement and agreement, and coordinate our actions. Supporting these interpersonal phenomena is plausibly a central role that language plays. Of course we can use language other ways (telling jokes, stories, etc.) but it will be enough for present purposes to attend to the more familiar interpersonal uses of language. This picture of language, as primarily a conventional tool for interpersonal coordination, lines up with Lewis’s view from his early book ‘Convention’ (1969) and several later papers (Lewis 1975; 1979; 1992; 1997).

What does linguistic content have to be like if language is going to support these coordinative phenomena? For one thing it must be robust in the face of deviant uses of expressions. A case adapted from Burge (1979) will serve as an illustration. Suppose Freda is suffering an intermittent pain in her thigh and visits her GP for medical advice. She is a basketball player and has heard that basketball players are susceptible to a condition called ‘arthritis’ in their limbs. She says to her doctor ‘I think I have arthritis in my thigh’. Freda is confused about arthritis and how ‘arthritis’ is used. She is, in a sense, using the term in a deviant way. But crucially Freda’s deviant use does not mean that the doctor cannot understand or get information from her utterance. Freda can use the term ‘arthritis’ successfully (such that it means what it means in the wider community) while being confused about how it is used.

Or consider an example borrowed from Bolinger (2020: 12).

Suppose an elderly neighbor volunteers hiring advice for your small business, saying “You’ve got to find yourself a Chinaman or a Jap. Ori-

---

<sup>1</sup>Some take ‘beliefs’ to be a kind of spurious plural in this context. We might have a holistic conception of belief such that the important conception of belief applies to a whole system of belief and not to small-scale discrete beliefs (Lewis 1986: 32). Though I have some sympathy with this view, it will not make a difference to the present discussion, since even die-hard holists about belief will be able to point to a *derivative* notion of belief such that each system of belief brings with it many beliefs in the derivative sense (Lewis 1986: 32).

entals have an incredible work ethic!” You thank her for the suggestion, but ask her not to use those particular terms in the future; they have become derogatory. To your dismay, she replies “Oh, I know they’re not politically correct anymore; people think they’re ‘offensive’. But that’s just what they’re called, and you know I don’t mean anything rude by it. People shouldn’t get so upset; there’s nothing actually bad about the words.” Suppose your neighbor is sincere, and does not (even implicitly) have contempt for people of Asian descent.

Let’s also suppose that she does not intend to derogate. For our purposes this is important since it is possible to exploit conventions to communicate even when the speaker does not have beliefs that line up with those conventions. For instance, see Stalnaker (2002: 716). When the neighbour uses those slurs, they have a derogatory meaning *despite* her sincerity, lack of the intention to derogate, and lack of contempt for the people of Asian descent. The meaning of the slur is insensitive to the intentions and beliefs of some individual users. This is what Bolinger calls ‘Derogatory Autonomy’ according to which ‘[w]hether a term is derogatory is largely independent of speaker intent. Sincerely not intending to derogate does not block the intimation of contempt, etc., involved in using a slur.’ (Bolinger 2020: 13) The impact of those terms is derogatory partly because of how the term is used in the wider linguistic community and its being derogatory is, as we have seen, insensitive to certain individual attitudes and intentions.

These examples bring out that linguistic content is insensitive to at least some idiosyncrasies of individual users. What an agent says when they make an utterance can come apart from what they take themselves to be saying or intend to say. This kind of robustness is plausibly part of why language can facilitate communication, disagreement, etc. Freda’s mistaken beliefs about arthritis and the term ‘arthritis’ do not threaten the claim that ‘arthritis’ means an inflammation of the joints, even when Freda uses the term. If we deny this kind of insensitivity, isolated deviant use of an expression would threaten to unduly undermine or change the meaning of a term in a public language, making public meaning too unstable to play its role in facilitating interpersonal communication and coordination.<sup>2</sup> The requirement that linguistic content be robust in this way might be characterised as part of what is required for linguistic content to be ‘public’. It is required for us to be *public speakers* when we use language.

## 2.2 Belief Content

Next up, beliefs. What are beliefs for? One important thing beliefs do is guide and explain reasoning and actions. When thinking and doing things in the world, we are often guided by how we take the world to be. If I believe that the faculty meeting is in seminar room A (and I want to attend the meeting), I will walk to seminar room A when the time comes. In a slogan borrowed from Ramsey (1929: 238), ‘a belief is a map of neighbouring space by which we steer’. This ‘space’ is generally understood as a vast multi-dimensional space with properties as locations so that our ‘maps’

---

<sup>2</sup>Of course linguistic content should not be understood as entirely insensitive to deviant uses of expressions, if a deviant pattern of use becomes widespread enough, it might be reflected in public meaning.

may guide not merely motion but cognition and behaviour more broadly. This idea has been largely taken up in the literature, though it is controversial how central this role for beliefs is. Certainly Stalnaker (1984), Lewis (1986: 34-40), and the many who follow them have emphasised this role as one of the most important of belief's roles.

If beliefs are to play this role, our characterisation of belief *content* should reflect this role. In particular, it should be sensitive to some of the believer's errors, ignorance, and idiosyncrasies.

Lois is looking for Superman when she walks past Clark's desk. What explains why she walks past Clark's desk? Part of the explanation seems to be that she does not believe that Clark is Superman. This seems part of the correct explanation of her behaviour even if it is widely known in her community that Clark is Superman. When it comes to explaining behaviour and the role that Lois' beliefs play in guiding her behaviour, the fact that she is ignorant of some fact should be reflected in our characterisation of the content of her beliefs about Clark.

Or consider again the case of Freda. Unless she is corrected, she will think, talk, and behave in accordance with her mistaken conception of arthritis. *How* her beliefs guide and explain her behaviour and reasoning will often be sensitive to her mistaken beliefs about arthritis and how the term 'arthritis' is used. How she *should* behave and reason (given the way the term is used in her community) is one thing, how she *does* act and reason given her mistaken beliefs is another, and if we are interested in explaining how she acts and reasons it is the latter that should concern us.

How a given belief guides an agent's action is also sensitive to her *other* beliefs. How my belief that Bronwyn is in her office will guide my actions will be sensitive to my other beliefs. Suppose I am a detective trying to hunt down a masked vigilante, 'the Kookaburra Queen', and I believe that Bronwyn is the Kookaburra Queen; the belief that Bronwyn is in her office will guide me in a very decisive way! I will rush to her office to arrest her! But that belief will not guide my action in that way if I do not believe that she is the Kookaburra Queen, it may not effect my behaviour in any significant way. So how a belief guides action depends on how it is integrated into a broader system of belief. So since *ex hypothesi* belief content is sensitive to how a given belief guides behaviour, we have reason to think that belief content is more holistic; the content of a given belief is partially determined by its relation to other beliefs. Indeed, prominent defenders of the conception of belief content discussed here insist on this kind of holism (Lewis 1986: 32; Braddon-Mitchell and Jackson 2007; Ramsey 1931). Though we don't express all our beliefs when we make an utterance, the content of the belief we express is plausibly sensitive to many of our other beliefs.

The content of beliefs, qua maps by which believers steer, seems to be sensitive to individual errors and ignorance. So belief content should be understood as being sensitive to the believer's individual psychological state, narrowly construed (Lewis 1986: 27-40; 1981). When drawing up, correcting, and being guided by our maps of the world, we are *on our own*. Facts about how others think and talk are only of use to us when they are reflected in our narrowly construed psychology. We are private investigators; though of course we can (and often do) gather information by interrogating witnesses.

From the discussion so far, we can already see that there is a kind of miss-match

between constraints on the contents of beliefs and the constraints on the content of linguistic expressions; the facts to which linguistic content should be insensitive seem to be precisely the facts to which belief content should be sensitive. Here I echo some of the remarks from Lewis (1986: 27-31), Braddon-Mitchell and Jackson (2007), Crimmins (1992), Heck (2002), Loar (1998), and others.

### 3 Autonomism

It seems like belief content is sensitive to believers' idiosyncrasies and linguistic content is often insensitive to the very same idiosyncrasies. For recent discussion, see Onofri (2016), Pollock (2015), Abreu Zavaleta (2020, 2021). One natural lesson we might take from this observation is that linguistic content and mental content are *autonomous*. Two agents could make the same assertion in the same context *and mean the same thing by their words* and yet be thereby expressing beliefs with different contents. Two agents could be expressing a belief with the same content while their utterances have quite different contents since they are situated in different linguistic communities. If we take the two kinds of content to be autonomous, both sets of constraints on content discussed above can have their way, so to speak. We are private investigators but public speakers. One might believe that the contents are autonomous *for other reasons*, but I reserve 'autonomism' for the claim that the contents are autonomous, motivated in the way just described.

### 4 Non-autonomist Options

The tension between constraints discussed above which autonomism dissolves has not gone unnoticed. The usual response is to ignore or weaken the lessons about mental and linguistic content; either by claiming that the contents of beliefs are insensitive to individual ignorance and error (belief contents are, in some sense, public), or by claiming that linguistic content is sensitive to individual errors and ignorance (linguistic content is, in some sense, private). The former is by far the most common response, though the latter is suggested by Frege (1918: 297) and has been discussed more recently by Heck (2012).

The idea that belief contents are public tends to go along with the view that the contents of beliefs often *just are* the contents of linguistic utterances.

The objects or contents of belief are also the objects or contents of speech acts, and it is the central task of semantics in general to say how utterances are related to the objects or contents of the speech acts they are used to perform. (Stalnaker 1988: 151)

If this is right, the content of an utterance is typically identical to the content of the belief that a speaker expresses when she makes that utterance.

There are a number of fairly familiar ways to secure this correspondence between mental and linguistic content. The view that thoughts are public is often associated with Frege (1892/1980; 1918-19/1956) and has been recently defended by Onofri (2018). How is the publicity of belief secured? There are various options here, for example one might appeal to knowledge of co-reference (Frege 1918: 297; Onofri

2018), de jure co-reference (Schroeter 2005; 2008; 2012; 2014; Schroeter and Schroeter 2009), or deferential uses of concepts (Sainsbury and Tye 2011; 2012).

If the content of beliefs is often shared, then it will need to be insensitive to individual differences in individuals' narrowly construed psychology. This gives rise to the notorious tension between the idea that thoughts are public and the idea that cases of fine-grained ignorance and error, and correspondingly fine-grained behavioural explanation, force us into a fine-grained conception of mental content. For recent illuminating discussions of this tension see Bjerring and Schwarz (2017), and Onofri (2018).

## 5 The Best of Both Worlds

One central advantage of autonomism is that it allows us to dissolve this tension and incorporate core insights from two influential traditions in the literature. Firstly, from the tradition associated with Donnellan (1970), Putnam (1975), and Kripke (1980), autonomists can take the insight that the meaning of expressions in a public language (including names and kind terms) is determined in part by things that need not be reflected in the narrowly construed psychology of individuals; '[m]eanings just ain't in the head!' (Putnam 1975: 227) This idea has been extremely influential and lines up naturally with the idea that linguistic content is insensitive to variation across individual language users.

But the autonomist can also incorporate the insights from a second tradition which is less popular nowadays, but is, in my view, no less important. Insights can be drawn from the work of Ramsey (1931), Lewis (1974; 1986: 34-40), Crimmins (1992), Loar (1998), Heck (2002), and Braddon-Mitchell and Jackson (2007). The central idea is that belief contents should be individuated finely because it allows beliefs to play the role they apparently play in fine-grained explanations of behaviour and cognition.

The central contentions within these traditions are often taken to be at odds with each other.<sup>3</sup> But those working in these traditions needn't fight, at least they needn't fight as much as they have been. Autonomism allows us to take central insights from both camps.

Middle ground positions tend to draw fire from both sides. The autonomist will likely be accused of being an agent of the Kripkean revolution while at the same time being vilified as a counterrevolutionary by those who welcome that same revolution. But autonomism is worth taking seriously, not least because it allows us to harmonise and respect these two important insights often taken to be at odds.

## 6 Pluralism, Hybridism, and Autonomism

It will be useful at this point to distinguish autonomism from two views, namely 'content pluralism' and 'content hybridism'. Autonomism is compatible with various kinds of pluralism about mental or linguistic content. For instance, Lewis (1986: 41-42 n.31) also suggests a kind of pluralism about the content of linguistic items,

<sup>3</sup>This tension is spelled out and interrogated by Onofri in an excellent series of papers (2016; 2018; 2019).

distinguishing what a name refers to ('David Lewis' refers to David Lewis) from its semantic value (which may be, say, a bundle of properties). Chalmers (2002; 2011) famously defends a pluralist conception of both linguistic content and mental content. There is nothing in pluralism per se that conflicts with autonomism, though some pluralists deny autonomism (see, for example, Chalmers (2011: 603)).

Another possibility is that autonomism undermines some of the *motivations* for pluralism. For instance, Chalmers (2002: 631 n.29; 2011: 620) claims that the reasons to ascribe secondary content to beliefs have to do with language, communication, agreement, disagreement, and the coordination of action and that to explain these phenomena we need a kind of mental content that is often shared across agents (and primary contents are not often shared). But if autonomism is correct, these phenomena could be explained in terms of public linguistic content rather than in terms of some shared dimension of mental content. This undermines the motivation for a Chalmers-style pluralism about mental content since explaining these interpersonal phenomena is the primary stated motivation for positing a second kind of belief content.

Then there is what we might call a *hybrid* conception of content (Shanklin 2017). Perhaps linguistic representations and mental representations have the same contents but their contents have two *aspects* such that the semantic value of linguistic items is captured by one aspect and the content of beliefs is captured in terms of the other. If this is right, strictly speaking, beliefs and sentences often have the same content and it is two aspects of the same content that account for linguistic and mental representation respectively. Aside from bookkeeping issues about how to use the word 'content', this kind of hybridism is compatible with autonomism. The autonomist need merely insist that these different aspects are themselves autonomous.

## 7 Overlap vs Decoupling

Members of a linguistic community will seldom have exactly the same beliefs about the meaning of an expression or what the expression picks out. They will also typically have different dispositions to use the expression. How should the autonomist understand the *public* meaning of an expression given this variation? In particular, does the mental content of the individuals in the community *determine* the content of the terms they use, and if so how? This is an important choice point for the autonomist.

One common idea is that linguistic content and understanding is determined by overlap in mental content among speakers of the language (e.g. beliefs about what a term means). This view is defended by Bezuidenhout (1997), Block (1998), Jackson (1998: 214), Prinz (2002: 159) and others. Let's call this the 'overlap view'.

The autonomist need not adopt the overlap view. Lewis (1997: 340-342), for one, contends that communication (and related phenomena) are possible even if there were no overlap between the mental content of individuals in a community. Lewis (1997: 340-342) observes that if linguistic content is tied to language's role as a tool that allows us to communicate, disagree, coordinate our actions, etc. without having any elements of mental representations in common. Lewis suggests an illustration, clearly inspired by Wittgenstein. Imagine two agents each with a vast

array of beetles in boxes (corresponding to their mental representations) but with no beetles in common and each agent can only look into their own boxes. These agents would, it seems, not be hampered in constructing and using rational systems of linguistic coordination between themselves (Lewis 1997: 342). They can indicate the boxes to each other and construct signalling systems without having to see into the other's boxes. Similarly, even if mental representations are not shared between agents, signalling systems for the transmission of information (in some sense of 'information') may arise between them. The same goes for larger communities, even if all Italian speakers have their own private mental representations and there is no overlap among their mental representations, the enormously complex signalling system that is Italian could still function. Correspondingly, the content of linguistic utterances might be common among the speakers despite the lack of mental overlap. A world where there is no mental overlap among humans appears to be a distant world; humans probably do have relevantly similar mental contents. Lewis's thought is that though there are similarities between the content of our mental representations, linguistic coordination does not, in general, depend on these similarities. Let's call this the 'decoupling view'. The decoupling view is compatible with the idea that overlap of mental content might be what determines linguistic content *in some cases*. The defender of the decoupling view will insist that shared linguistic content is not always determined by similarities between mental contents.<sup>4</sup>

Which view should the autonomist adopt here? I am more sympathetic to the decoupling view. For one thing, it better reflects what I take to be the primary role of language as a tool for coordination. I also find some recent arguments presented by Onofri (2019) against the overlap view (which he calls the 'Similar Ways of Thinking' view) convincing. In any case, we can stay largely neutral on this issue for present purposes; autonomism does not necessarily stand and fall with the overlap view.

## 8 The Interface

The autonomist sees distinctions between belief contents where others see none. This conception of belief content has some apparent costs. When compared with some of her non-autonomist rivals, the autonomist is forced into adopting a comparatively messy view of assertion, communication, disagreement: phenomena, in short, which involve the *interface* between linguistic and mental content. This messiness may give rise to a number of objections; these objections will be the primary target of the rest of the paper.

I have already discussed assertion. A very natural account of assertion is that we use assertions to express our beliefs and to express a belief is to make an assertion whose content just is the content of the belief. If autonomism is correct, this simple view of assertions will have to be revised.

Next consider communication. Suppose I believe that Bronwyn is in her office and you ask 'is Bronwyn at work today?' I then produce the utterance 'Bronwyn is in her office' to both express my belief and answer your question. You trust me

---

<sup>4</sup>At least when similarities are understood in the normal way, in terms of some kind of discrete element of content in common.



and understand what I said and come to believe that Bronwyn is in her office. An elegant and plausible model of communication says that successful communication like this involves an agent expressing a belief with an assertion and the audience understanding the utterance and forming a belief with the same content as the belief the speaker expressed (Stalnaker 1978; 1988, 151). Borrowing some terminology from Weber (2013: 207-208), let's call this the *FedEx* model of communication. If autonomism is right and belief content is sensitive to individual ignorance and error, then it will seldom be the case that the belief I express and the belief you form will have exactly the same content. Maybe I am like Lois and am unaware of Bronwyn's masked vigilante alter ego while you are in on the secret (and sworn to secrecy). If the content of my beliefs about Bronwyn is sensitive to these differences then the content of our beliefs concerning Bronwyn will be different. So, according to the FedEx model, we did not successfully communicate. If the FedEx model is correct, then the claim that belief content is sensitive to individual ignorance and error will imply that communication is rare. Since communication is not rare, either the FedEx model or autonomism must go.

Next consider disagreement. Suppose Peter has just seen Bronwyn leaving her office for morning tea and overhears our exchange. Peter believes that Bronwyn is not in her office and cuts in saying 'Bronwyn is not in her office'. A natural thought is that this is a point of genuine disagreement because there is some one proposition (which captures a crucial element of the content of our beliefs) that one of us rejects and the other accepts (in this case the proposition is plausibly 'Bronwyn is in her office'). It is also plausible that there is a kind of *conflict* between my belief and Peter's which obtains even if our disagreement about how things are never comes out in what we say. This belief conflict is also plausibly captured in terms of there being some proposition that one of us believes that the other rejects. Huvenes (2012: 178-179) gives a similar story about conflicting attitudes other than belief, for example desires and preferences can conflict, even if the conflict does not come up in conversation. The idea is that when there is a conflict, there is some content that is shared between the parties to the disagreement. Given the almost undeniable claim that genuine disagreement is common, this simple model of disagreement is incompatible with autonomism. Abreu Zavaleta (2020) has recently argued convincingly against this simple model of disagreement. If Abreu Zavaleta's claims are correct, and I am inclined to think they are, then the incompatibility of autonomism and this simple conception of disagreement would be no reason to reject autonomism.

Next consider agreement. We often have beliefs which agree and I am often thinking what you are thinking. When two beliefs agree, they stand and fall together; it cannot be (in some sense of 'cannot') that one belief is true while the other is false. Suppose we both believe that Dobbin is a horse. Again a natural account of how beliefs might stand and fall together is that their content is the same and is captured by the same proposition. This model of agreement is unavailable to the autonomist. There will almost always be differences between individuals' ignorance and error about that object, leading to differences in which proposition captures the content of different agents' beliefs. More carefully, we will seldom be able to find a proposition that captures the content of both beliefs if propositions are understood in the orthodox way, i.e. as sets of worlds. In turn, given that agreement is common, either autonomism or the simple model of agreement will

have to go.

## 8.1 The Debt

So there are a number of models of communication, assertion, disagreement, and agreement which are unavailable to the autonomist. Does this mean that we should reject autonomism? I don't think so, at least not immediately.

The fact that autonomism is incompatible with elegant models of these phenomena means that autonomists incur a kind of explanatory debt. They must develop alternative autonomist-friendly models with comparable predictions and power. Autonomists should take this debt seriously and avoid defaulting if they can. That being said, the independent motivations for autonomism mean that it is worth exploring whether and how this debt might be paid.

## 8.2 Payment Plans

There are a number of promising payment plans open to the autonomist. Attitudes have many features that do not supervene on their content and stand in relations to the world and each other aside from the 'has the same content as' relation. Perhaps one of these relations might serve as a basis for alternative accounts of communication, disagreement, and the rest. I cannot hope to fully develop an autonomist-friendly account of these phenomena here, though I will attempt to give an outline of some promising options.

What other relations could do the job of underpinning phenomena at the interface? One natural idea is that the 'about the same thing' relation might do the job. For example, maybe my belief that Bronwyn is in her office and Peter's belief that she is not conflict partly because both are about Bronwyn.<sup>5</sup> To see how this might work, let's start with Lewis's (1981) autonomist-friendly conception of attitude content from 'What puzzling Pierre does not believe'. Suppose Pierre has the belief he would express by saying 'London is not pretty' and the belief that he would express by saying that 'Londres is pretty'. I have simplified the case from Kripke's (1979) original, abstracting away from issues about translation. According to Lewis, the contents of Pierre's beliefs are that there is something that plays the London role that is not pretty and something that plays the Londres role that is pretty. These 'roles' are sensitive to his ignorance and error, and therefore end up being extremely fine-grained.<sup>6</sup> These contents are compatible, there is a possible

---

<sup>5</sup>Does this mean understanding the content of the belief as being at least partly referential? Not necessarily. Some prominent defenders of fine-grained belief content and those who follow them (Chalmers 2002: 631 n.29; 2011: 602, 620–621; Kipper 2018) argue that there is also a referential element to mental content which supports disagreement and communication. According to these proposals, there is a coarse-grained referential element to attitude content, the identity of which supports phenomena at the interface. But autonomists need not think that what an attitude is about is reflected in its *content*. The autonomist can plausibly deny that the aboutness of a representation is a kind of content (or determines a kind of content), see below. I have also argued elsewhere that, in any case, referential dimensions of content do not explain disagreement and communication (Sandgren 2018).

<sup>6</sup>Lewis also claims that these belief contents are fine-grained because they are egocentric contents (Lewis 1979). For a good recent discussion and clear formulation of Lewis's view on de se beliefs, as well as strong replies to its critics, see Openshaw (2020)

situation in which they are both true (e.g. a world in which one pretty thing plays the Londres role and a different non-pretty thing plays the London role).

But there *is* a sense in which these beliefs conflict. If these were beliefs of two different agents, we would naturally say that they disagree about London. Here I echo some remarks from my 'Puzzling Pierre and Intentional Identity' (Sandgren 2019: section 3). How can we account for this conflict without coarse-graining belief content? Lewis's discussion of Kripke's puzzle about belief in 'On the Plurality of Worlds' (1986: 33) suggests a way. Even if the content of Pierre's belief does not include London, his belief is *about* London. One way to explain this aboutness is in terms of his being connected to London via two causal chains of acquaintance. These causal chains can be more or less long and the corresponding acquaintance may be more or less direct (Lewis 1986: 33). For instance one might have seen London for oneself or one might have only read about it in a book. Perhaps the beliefs are plausibly about the same thing since the two causal acquaintance relations lead to the same city. For the purposes of outlining autonomist-friendly accounts of phenomena at the interface, we can work with this kind of simple causal account. Though I should emphasise that there are other ways of understanding aboutness and subject matter in general that are not purely causal, more on this below.

The autonomist can then contend that the content of a belief is one thing and its subject matter (what it is about) is another. For a discussion of this distinction see Lewis (1986: 33) and some more recent work by Jackson (2015), Sandgren (2019), and Pepp (2019). and that belief content is fine-grained and seldom shared while subject matter is coarse grained and often shared.<sup>7</sup> It is then open to the autonomist to claim that it is shared subject matter that is the arbiter of genuine disagreement; the relation of being about the same thing could then plausibly explain (or help to explain) the conflict between Pierre's beliefs. They conflict because they are about the same thing and they ascribe conflicting properties.

On this account Peter disagrees with me in the case described above because we are both thinking and talking about Bronwyn and we ascribe incompatible properties to her. This conception of disagreement is compatible with the claim that there is no element of content corresponding to Bronwyn that forms part of the content of our respective beliefs.

There are also parallel possibilities for aboutness-based accounts of agreement, communication, and assertion. To start with communication. Suppose the content of the belief expressed by the speaker and the belief formed by the audience have different contents (and that there is no discrete element of content in common across those belief). They may yet be communicating if the beliefs are about the same things. When I say 'Bronwyn is in her office' and you form a belief on the basis of what I said, we could say that the communicative interaction is successful only when the belief expressed and the belief formed are about the same things. Properties are often part of a belief's subject matter (in this case 'being in her office') and for communication to be successful in this case this property must be one of the things the two beliefs are about. If you end up with a belief about a totally different

---

<sup>7</sup>This aboutness should be distinguished from the kind of aboutness Yablo (2014), Fine (2014: 2016), and others discuss. For Yablo and Fine, what a representation is about depends on what makes it true or what would make it true. On the contrary, according to the suggestion here the aboutness of a belief does not supervene on its content or truth conditions.

property and ascribe it to Bronwyn, then communication may have failed for that reason.

Next, agreement. Suppose you and I believe that Dobbin is a horse. There is a sense in which our beliefs stand and fall together. But if the autonomist is right, this will not be because they have the same content. Instead we can say that *because* they are about the same thing and property, they will stand and fall together in the sense that one cannot be true while the other is false *without changing what they are about*. If this is right, to think 'the same thing' is not to have beliefs with the same truth conditions, rather it is, at least in part, a matter of beliefs having the same subject matter (objects and properties).

Finally, let's turn to assertion. Linguistic items also have subject matter. The sentence 'Jill is ill' is about Jill (and being ill). This observation suggests a natural account of the relationship between assertions and the beliefs they express. When I say 'Jill is ill' the belief I am expressing will not have the same content as the assertion itself, according to the autonomist. But the autonomist can say that part of what makes for successful assertion (and plays many of the explanatory roles content identity is typically taken to play) is that the belief expressed and the sentence uttered have the same subject matter, namely Jill and being ill.

Of course this is only a brief sketch of autonomist-friendly accounts of these phenomena at the interface. A number of the crucial notions involved will need to be unpacked and their details spelled out. Most importantly, aboutness (and sameness of aboutness) has been largely treated as a black box here. I briefly sketched one way of understanding subject matter, as being determined by relations of causal acquaintance with objects in the world. But purely causal ways of making sense of subject matter and sameness of subject matter are crucially limited (Sandgren 2019; 2021). To begin with, purely causal accounts of aboutness cannot handle representations of abstracta (e.g. numbers) and intentional targets that appear not to exist (e.g. witches), since, in both cases, the respective subject matters do not seem to stand in the right sort of causal relations of acquaintance with believers. It is, however, extremely plausible that we can make assertions, communicate, agree, and disagree about abstracta and intentional targets that, in some sense, turn out not to exist. So, at best, causal relations of acquaintance are only part of the story about aboutness. These considerations put some constraints on the theory of aboutness that autonomists ought to plug into the proposed autonomist-friendly accounts of phenomena at the interface. Of course, there are many important questions about the nature of this kind of aboutness that must be left open here.

However, there is one feature of aboutness, as I am conceiving it, that bears emphasis at this point. The aboutness of a belief should be distinguished from what we might call the 'reference' of a representation. A belief about Stockholm might refer to an object in a way that a belief about the fountain of youth does not (since there is no fountain of youth). Yet apparently both beliefs can have aboutness; it seems as if we can think about Stockholm and we can think about the fountain of youth. Of course there are those who argue that reference and aboutness correspond in these cases, either because we cannot really think about the fountain of youth or because there really is, in some sense, a fountain of youth that the relevant beliefs refer to. But, as I conceive of them, reference and aboutness are distinct features of representations.

For aboutness-based accounts of phenomena at the interface to work, we will

need to plug in an adequate account of aboutness. It is beyond the scope of this paper to compare accounts of aboutness or develop a full picture of aboutness, qua element of the autonomist explanations of assertion, communication, agreement, and disagreement I sketched above. Nonetheless, the above observations about cases that are trouble for simply causal views suggest that whichever account we plug in, it had better be generalisable to these cases. There are a number of available accounts of aboutness and co-aboutness that fulfill this criterion. For instance, there is the concept-based approach of Sainsbury (2010), the phenomenological approaches defended by Farkas (2008), and Mendelovici (2018), the aspect-based account due to Geach (1976), the co-aboutness first approach due to Sandgren (2021), and the cognitive role-based accounts due to Jeshion (2002: 67), Azzouni (2011), and Crane (2011), to mention a few.

There is also an open question about how properties form part of the subject matter of representations that I haven't answered here. But I hope that I have done enough to suggest one promising way the autonomist could attempt to pay off their explanatory debt; aboutness (and relations of being about the same thing) may be able to play many of the roles traditionally attributed to content and identity (or similarity) of content across representations.

Another promising option that I should mention is based on Weber's 'Recentering Model of Communication' (2012: 210-213). The idea is, roughly, that when we hear and understand utterances we extract information from the utterance in accordance with our beliefs about how the speaker is related to ourselves. This means that even if the contents of the belief expressed and the belief formed are different (and have no element of content in common) we can still *properly* extract information from utterances in a way that constitutes successful communication. Weber (2012: 216-218) argues that disagreement and agreement might be captured with a parallel 'Recentering' model of disagreement and agreement which does not require belief contents to be shared across the relevant believers. Weber also contends that the model can be extended in a natural way to account for the content of assertion, at least in simple cases (Weber 2012: 281-282). Jackson (2010: section 7) presents a similar account of communication and assertion based on centered worlds. Of course, more work is required to explore whether models of this kind capture the features of communication, disagreement, etc. that they are supposed to capture. It is also unclear whether the recentering models should be understood as rivals to or supplementary to the aboutness-based models described above. I am inclined to think of the proposals as supplementing each other, but more work is needed to understand if and how they fit together.

The last move I want to discuss involves a kind of Davidsonian (1973; 1974; 1985) appeal to charity. The autonomist could argue that messy beliefs and systems of beliefs don't interface with language directly. Instead, tidied up proxies for beliefs, the result of applying certain principles of charity to the beliefs and their contents, do. When beliefs interface with language, we could apply principles of charity to the relevant beliefs and their contents, thereby producing more coherent, reasonable, sensitive to the evidence, etc., rational reconstructions or 'proxies' for the beliefs and their contents which fit more easily with public language. The hope would be that this proxy would be more in line with the determinants of the content of a public language and would, therefore, help bridge the gap between mental and linguistic content. In turn, the autonomist could claim that the relations

between beliefs that support communication, assertion and the rest are tied to how the charitably reconstructed analogs of beliefs are related. For example, two such proxies for distinct beliefs (with distinct contents) might have the same or similar content. Again, many details of this move would have to be filled in here, and many crucial questions answered (what does this charity amount to?, can these tidied up proxy contents really interface cleanly with language?, etc.). However, it is an interesting option worthy of exploration.

There are a number of promising possibilities for paying the debt without compromising autonomism or the motivations for autonomism (I have only been able to briefly outline a handful), so I don't think the autonomist should resign themselves to defaulting just yet.

### 8.3 Belief Reports

Autonomism also requires a complicated account of the relationship between the content of attitudes and attitude reports. Again, this is not surprising. Attitude reports are linguistic entities and their subject matter is, at least partly, the content of belief. Like assertion and the rest, attitude reports involve the interface between mental content and linguistic content.

There is a conception of the relationship between belief content and belief reports according to which:

'Pierre believes that  $F(A)$ ', where  $A$  is an ordinary proper name and  $F$  is an easily understood predicate, ascribes to Pierre a belief whose object is the proposition (actually) expressed by ' $F(A)$ '. (Lewis 1981: 284)

On this kind of story, 'believes that ...' constructions have the role of expressing the proposition which is the content of the ascribed belief. We can *read off* the content of beliefs from the content of belief reports in a straightforward way. Let's call this the 'carbon copy' (CC) theory of attitude reports. The CC theory also supports a simple disquotation principle along the following lines:

'Pierre believes that London is pretty' is true just in case Pierre believes that London is pretty.

The CC theory is a simple and plausible view.

[The CC theory] surely deserves to be held, but for its failures: it is simple and plausible, and it fits perfectly into a systematic program for compositional semantics. If it did not exist, it would be necessary to invent it – and to refute it. (Lewis 1981: 285)

The autonomist must resist the CC theory. Is this a heavy cost? I don't think so. Firstly, one might be driven to reject the CC theory anyway. For instance, Lewis (1981: 283-285) convincingly argues that Kripke's (1979) puzzle about belief provides grounds for a refutation of the CC theory. Roughly, Lewis contends that Kripke's puzzling case is one in which we can arrive at a false view about an agent's beliefs (that they believe an obvious contradiction) *if* we accept the CC theory of belief reports, and it is the CC theory that should be jettisoned. I take arguments like these to weigh heavily against the CC theory. Lewis (1981: 285-287) and Jackson

(2015) also give more general *epistemic* objections to the CC theory according to which the CC theory implies that believers have significant knowledge about the *essences* of the object their beliefs are about when they do not. I also find these arguments convincing and if we have independent reason to reject the CC theory, it is no cost that there is a tension between autonomism and the CC theory.

Secondly, the autonomist has tools ready to hand to build an alternative approach to belief reports (Lewis 1986: 32-33). Suppose Ralph is watching a spy at work in the shadows but does not recognise that the spy is Bernard. 'Ralph believes that Bernard is a spy' can be true in this case, even if Ralph does not realise, believe, or know that the spy is Bernard. The autonomist can say that 'Ralph believes that Bernard is a spy' has a mixed subject matter in that it is not only about the content of his beliefs (which we may take as being as the autonomist says they are), but also about which objects what his beliefs are *really* about. As Lewis (1986: 33) suggests, 'Bernard gets into the act... because he is the one Ralph is actually watching.' The fact that belief reports concern these facts about what the belief is about that may not be reflected in the believer's narrowly construed psychological state does not imply that those same facts make a difference to the content of the relevant beliefs. The report says that he has a belief that there is a spy at work in the shadows, and that that belief really is about Bernard.

The so-called belief sentence "Ralph believes that Bernard is a spy" has a mixed subject-matter. It is not entirely about Ralph's system of belief. It is made true partly by Ralph's psychological state, and partly by his relationship to his surroundings. (Lewis 1986: 33 n.25)

We can then say that

'Ralph believes that Bernard is a spy' (when read as a *de re* attitude report) is true just in case Ralph has a belief with content *c* and that belief is really about Bernard (and spyhood).

In this case, the latter part of the right-hand-side of the biconditional is not reflected in Ralph's narrowly construed psychology (he does not recognise Bernard). This 'mixed subject matter' model is natural, attractive, and consistent with autonomism. Like some of the other accounts of phenomena at the interface sketched above, this model requires a non-trivial distinction between a belief's content and what it is about along the lines of that discussed in the previous section.

Does giving up on the CC model mean giving up on a systematic program for compositional semantics? It is clear that we will have to give up on the *kind* of compositionality supported by the CC theory and simple disquotation principles. As one would expect, many forms of compositionality will be more difficult to ensure given autonomism. Autonomism implies a mess at the interface and attitude reports are at the interface. But there is reason to believe that we could build a compositional semantics for belief reports in line with the mixed subject matter model. Also, if we go with the aboutness-based proposal, whether we can establish a compositional semantics will depend on whether aboutness is amenable to a compositional treatment. Again, I cannot fully spell out this or any other promising autonomist-friendly model of belief reports here, but I hope I have shown that there are promising and exploration-worthy options available for making sense of belief reports in an autonomist-friendly way.

But even if autonomism does imply that our account of belief reports will need to be messier than we might have liked, that is no reason to reject autonomism immediately. The study of attitudes and their contents is one thing, a study of attitude reports is another. Why should the elegance of a theory of attitude reports dictate our theory of attitude content? Why not reason the other way around? If the autonomist treatment of mental content is independently motivated, as I have argued it is, then there is some pressure to accept the consequences thereof. It is not obviously the right choice to make our conception of mental content unfit for purpose to make it suit a given theory of attitude reports. I am aware that this point is quite controversial; there are those who think that belief reports and their structure are one of our best guides to studying belief content, but those who are tempted by autonomism need not sign on to this kind of reports-first approach.

## 9 Taking Stock

Linguistic content and mental content seem to play different roles. These different roles appear to call for linguistic content and mental content to be sensitive to different things. The lesson the autonomist takes from this is that linguistic content and mental content should be taken to be *autonomous*. This option is seldom discussed in the contemporary literature. Most opt to compromise the role of belief contents in guiding behaviour and cognition, making belief contents coarser grained such that they are often shared across agents. Autonomism allows us to uphold the apparently conflicting constraints on mental and linguistic content without compromise. For this reason it is worth taking seriously, even if adopting autonomism forces us to develop relatively complex models of phenomena at the interface of mental and linguistic content.<sup>8</sup>

## References

- Abreu Zavaleta, M. (2020). Disagreement lost. *Synthese*, 1–34.
- Abreu Zavaleta, M. (2021). Communication and variance. *Topoi* 40(1), 147–169.
- Azzouni, J. (2015). Singular Thoughts (Objects-Directed Thoughts). *Aristotelian Society Supplementary Volume* 85(1), 45–61.
- Bezuidenhout, A. L. (1997). The communication of de re thoughts. *Noûs* 31(2), 197–225.
- Bjerring, J. C. and W. Schwarz (2017). Granularity problems. *Philosophical Quarterly* 67(266), 22–37.
- Block, N. (1998). Conceptual role semantics. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*, pp. 242–256. Routledge.

---

<sup>8</sup>I would like to thank Daniel Nolan, Martín Abreu Zavaleta, Jussi Haukioja, Manuel García-Carpintero, Frank Jackson, Pär Sunström, Torfinn Huvenes, and Andrea Onofri for feedback on the material in this paper. Special thanks to Renée Jorgensen, who first suggested the idea for this paper to me. This research was supported by a Vetenskapsrådet Research Project Grant (2019-02786).



- Bolinger, R. J. (2020). Contested slurs. *Grazer Philosophische Studien* 97(1), 11 – 30.
- Braddon-Mitchell, D. and F. Jackson (2007). *The Philosophy of Mind and Cognition*. Blackwell.
- Burge, T. (1979). Individualism and the mental. *Midwest Studies in Philosophy* 4(1), 73–122.
- Chalmers, D. (2002). The components of content. In D. J. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Chalmers, D. J. (2011). Propositions and attitude ascriptions: A Fregean account. *Noûs* 45(4), 595–639.
- Crane, T. (2015). The Singularity of Singular Thought. *Aristotelian Society Supplementary Volume* 85(1), 21–43.
- Crimmins, M. (1992). *Talk About Beliefs*. MIT Press.
- Davidson, D. (1973). Radical interpretation. *Dialectica* 27(1), 314–328.
- Davidson, D. (1974). Belief and the basis of meaning. *Synthese* 27(3-4), 309–323.
- Davidson, D. (1985). Incoherence and irrationality. *Dialectica* 39(4), 345–354.
- Donnellan, K. S. (1970). Proper names and identifying descriptions. *Synthese* 21(3-4), 335–358.
- Farkas, K. (2008). Phenomenal intentionality without compromise. *The Monist* 91(2), 273–293.
- Frege, G. (1892/1949). On sense and nominatum. In H. Fiegl and W. Sellars (Eds.), *Readings in Philosophical Analysis*, pp. 85–102. Appleton-Century-Crofts.
- Frege, G. (1918/1956). The thought: A logical inquiry. *Mind* 65(259), 289–311.
- Geach, P. T. (1976). Two kinds of intentionality? *The Monist* 59(July), 306–320.
- Heck, R. (2002). Do demonstratives have senses? *Philosophers' Imprint* 2, 1–33.
- Heck, R. (2012). Solving Frege's puzzle. *Journal of Philosophy* 109(1-2), 132–174.
- Huvenes, T. T. (2012). Varieties of disagreement and predicates of taste. *Australasian Journal of Philosophy* 90(1), 167–181.
- Jackson, F. (1998). Reference and description revisited. *Philosophical Perspectives* 12(S12), 201–218.
- Jackson, F. et al. (2015). Singular belief. In *Passions and Projections Themes from the Philosophy of Simon Blackburn*. Oxford University Press.
- Jeshion, R. (2002). Acquaintanceless de re belief'. In J. K. Campbell, M. O'Rourke, and D. Shier (Eds.), *Meaning and Truth: Investigations in Philosophical Semantics.*, pp. 53–74. Seven Bridges Press.

- Kipper, J. (2018). Communicating egocentric beliefs: Two-content accounts. *Erkenntnis* 83(5), 947–967.
- Kriegel, U. (2008). The dispensability of (merely) intentional objects. *Philosophical Studies* 141(1), 79–95.
- Kripke, S. A. (1979). A puzzle about belief. In A. Margalit (Ed.), *Meaning and Use*, pp. 239–83. Reidel.
- Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review* 88(4), 513–543.
- Lewis, D. K. (1969). *Convention: A Philosophical Study*. Wiley-Blackwell.
- Lewis, D. K. (1974). Radical interpretation. *Synthese* 27(July-August), 331–344.
- Lewis, D. K. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota Studies in the Philosophy of Science*, pp. 3–35. University of Minnesota Press.
- Lewis, D. K. (1981). What puzzling pierre does not believe. *Australasian Journal of Philosophy* 59(3), 283–289.
- Lewis, D. K. (1986). *On the Plurality of Worlds*. Wiley-Blackwell.
- Lewis, D. K. (1992). Meaning without use: Reply to hawthorne. *Australasian Journal of Philosophy* 70(1), 106–110.
- Lewis, D. K. (1997). Naming the colours. *Australasian Journal of Philosophy* 75(3), 325–42.
- Loar, B. (1998). Is there a good epistemological argument against concept-externalism. *Philosophical Issues* 9, 213–217.
- Mendelovici, A. (2018). *The Phenomenal Basis of Intentionality*. New York, USA: Oxford University Press.
- Onofri, A. (2016). Two constraints on a theory of concepts. *Dialectica* 70(1), 3–27.
- Onofri, A. (2018). The publicity of thought. *Philosophical Quarterly* 68(272).
- Onofri, A. (2019). Loar’s puzzle, similarity, and knowledge of reference. *Manuscrito* 42(2), 1–45.
- Openshaw, J. (2020). Self-ascription and the de se. *Synthese* 197(5), 2039–2050.
- Pepp, J. (2019). Principles of acquaintance. In T. Raleigh and J. Knowles (Eds.), *Acquaintance: New Essays*. Oxford University Press.
- Pollock, J. (2015). Social externalism and the problem of communication. *Philosophical Studies* 172(12), 3229–3251.
- Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press.
- Putnam, H. (1975). The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* 7, 131–193.

- Ramsey, F. P. (1929). General propositions and causality. In F. P. Ramsey (Ed.), *The Foundations of Mathematics and other Logical Essays*, pp. 237–255. Kegan Paul, Trench, Trübner.
- Sainsbury, R. M. (2010). Intentionality without exotica. In R. Jeshion (Ed.), *New Essays on Singular Thought*.
- Sainsbury, R. M. and M. Tye (2011). An originalist theory of concepts. *Aristotelian Society Supplementary Volume* 85(1), 101–124.
- Sainsbury, R. M. and M. Tye (2012). *Seven Puzzles of Thought and How to Solve Them: An Originalist Theory of Concepts*. Oxford University Press.
- Salmon, N. U. (2005). *Metaphysics, Mathematics, and Meaning*. Oxford University Press.
- Sandgren, A. (2018). Secondary belief content, what is it good for? *Philosophical Studies* 175(6), 1467–1476.
- Sandgren, A. (2019a). A metarepresentational theory of intentional identity. *Synthese* 196(9), 3677–3695.
- Sandgren, A. (2019b). Puzzling pierre and intentional identity. *Erkenntnis* 84(4), 861–875.
- Sandgren, A. (2021). Turning aboutness about. *Pacific Philosophical Quarterly*.
- Schroeter, L. (2008). Why be an anti-individualist? *Philosophy and Phenomenological Research* 77(1), 105–141.
- Schroeter, L. (2012). Bootstrapping our way to samesaying. *Synthese* 189(1), 177–197.
- Schroeter, L. (2014). Normative concepts: A connectedness model. *Philosophers' Imprint* 14.
- Schroeter, L. and F. Schroeter (2009). A third way in metaethics. *Noûs* 43(1), 1–30.
- Shanklin, R. (2017). Local meaning, public offense. *ProtoSociology* 34, 163–177.
- Stalnaker, R. (1984). *Inquiry*. Cambridge University Press.
- Stalnaker, R. (1988). Belief attribution and context. In R. H. Grimm and D. D. Merrill (Eds.), *Contents of Thought*, pp. 140–156. University of Arizona Press.
- Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy* 25(5-6), 701–721.