

# *From externalism to expressivism*

BENJ HELLIE

CaSE: NYU, 12–13 May 2017

## 1 Outline

A dialectic from [Hellie 2011, 2013](#):

- The **conclusion** is that we should be (simulationist) **expressivists** about consciousness-discourse: that when we say this-and-that regarding what it is like for so-and-so, we put on display our ‘re-thinking’ ([Collingwood 1946/1993](#)) of so-and-so as involving this-and-that, without thereby dividing modal space;
- The **assumptions** behind the conclusion add up to a certain **externalism** about consciousness: first, a weakish *rationalism* regarding consciousness, to the effect that rationalization ‘supervenes on’ what it is like (in the nonprejudicial sense that *treating two subjects as the same* in what it is like requires *treating them as the same* in regard to rationalization); and second, that what rationalizes includes *evidence*—which, for familiar reasons, should be understood along broadly ‘externalist’ lines;
- The **argument** from the externalist assumptions to the expressivist conclusion runs through **hallucination**: externalists about evidence see hallucination as involving inconsistency in belief; but inconsistent belief involves ‘fragmentation’; and a discourse of fragmented consciousness makes sense only if it is expressive.

The bulk of the handout concerns the interior of the dialectic:

- (§3) Why we should think of evidence in a way that requires a fragmental treatment of hallucination
- (§4) How the expressivist treatment of fragmentation works

In the interest of pushing questions of detail off to the end, I begin with two ‘shell’ issues:

(§2.1) Why *rationalism*?

(§2.2) Why an *expressivist* interpretation of fragmental rationalism?

## 2 Shell issues

### 2.1 Rationalism—why?

*def* *Rationalizations* are ‘rationalizing explanations’—those where the explanandum is the fact that someone acquires a certain theoretical or practical attitude and the explanans is some other mental fact about them

⇒ Rationalizations (concerning a subject  $j$  at a time  $t$ ) supervene on what it is like (for  $j$  at  $t$ )

- This is weaker than the claim that rational-psychological facts supervene on what it is like (if there are ‘inert’ rat-psych facts, they are irrelevant); and it can be distinguished from its converse, which I set to the side
- One motivation for rationalism is a [Horgan and Tienson 2002](#)-esque ‘phenomenal intentionalism’; another goes by ‘four brains’-type reasoning ([Hellie 2007](#), 296); another goes by a general Collingwoodesque way of thinking about mental discourse ([Heal 2003](#); compare [Hellie 2017](#))

## 2.2 From fragmental rationalism to expressivism

- A good case–bad case pair:

Each of Good and Bad is instructed to pay attention to how things are for them visually: either a green light or a red light will be flashed; if they see a green light flash, they are to push ‘G’; if they do not see a green light flash, they are to push ‘R’.

For Good, a green light is flashed: Good then pushes ‘G’.

For Bad, a red light is flashed: Bad nevertheless pushes ‘G’. Why? — Unbeknownst to Bad, something weird was done to their visual cortex just over the period of the flash, to make it neurally the same as Good’s during the period of the flash.

- The *fragmental* approach to Good and Bad (Hellie 2011, 137, 143, 153; Hellie 2013, 132–3)
  - Begin with a broad ‘direct realism’, whereby Bad and Good differ in their evidence (Bad’s evidence has content *b* about which kind of bad case Bad inhabits, while Good’s evidence has incompatible content *g* about which kind of good case Good inhabits)
  - What leads Bad astray is a shared background presumption with Good about how to ‘interpret’ their evidence (each presumes they are in a good case), in consequence of which each interprets their evidence in the same way—both interpretations with the content *g* (Good correctly, Bad incorrectly)
  - But both evidence and its interpretation are believed: so Good believes a certain good-case proposition *g* both as evidence and as its interpretation; while Bad believes the bad-case proposition *b* as evidence and *g* as its interpretation
  - Unfortunately, *b* and *g* are incompatible, so Bad has inconsistent beliefs
  - And, following Lewis 1982, 103–4 (compare Elga and Rayo 2009; Rayo 2013, 4.3, 4.5), Bad is therefore in a ‘fragmented’ mental condition, with one fragment believing *b* (and therefore not-*g*), and another believing *g* (and therefore not-*b*)
- *Expressivists* about a discourse think it does not divide modal space:

- This allows for *faultless disagreement*: assertions within the discourse that are both correct, yet display incompatible mental states
- Accordingly, expressivism about consciousness allows for faultless disagreement how to treat so-and-so in regard to consciousness
- And Bad’s fragmentation can be accommodated as ‘faultless disagreement with oneself’ over whether to treat Bad with the *b*-fragment or with the *g*-fragment:
  - \* The *b*-fragment is what we find if we think our way into Bad’s mind starting with Bad’s visual evidence and reasoning forward;
  - \* The *g*-fragment is what we find if we start with Bad’s motor-activity evidence and reasoning backward

- But the *descriptivist* about consciousness thinks consciousness-discourse does divide modal space: accordingly, the world selects exactly one of the following implausible options regarding how Bad and Good are (relevantly, in respect of consciousness):
  - *Exactly the same*—but then the *b*-fragment is kicked out of what supervenes on consciousness, and evidence is generally ‘screened off’
  - *Entirely different*—but then the *g* fragment is kicked out of what supervenes on consciousness, and the central intuition over good–bad pairs is neglected
  - *The same in a respect, different in a respect*—this preserves both the *b* and the *g* fragment, does not ‘screen off’ evidence, and partly accommodates the good–bad intuition; but it does not fully accommodate the intuition; and it also makes for an unrecognizable characterization of what it is like: evidence is ‘laid on the slab’ alongside a distinct, clashing interpretation—eh?

### 3 Evidentialism

An alluring idea: perception somehow ‘exogenously’ feeds belief with facts—after all, that makes it OK to conditionalize on those propositions (whereas other sources of belief would be nonmonotonic)

⇒ We have ‘evidence’ consisting of believed truths arrived at somehow exogenously, through perception

*def* A subject  $j$  at a time  $t$  *treats-as-evident* a proposition  $e$ , to a subject  $j^*$  at a time  $t^*$ , **only if**  $j$ -at- $t$  believes (i)  $e$ ; and (ii) that  $j^*$ -at- $t^*$  treats  $e$  as evident to  $j^*$ -at- $t^*$

What else?

#### 3.1 Subject-matter evidentialism?

⇒ One has evidence whether  $\varphi$  just if whether  $\varphi$  concerns one’s ‘experience’ (Lewis 1972/1997, 1996)

##### Issues

1. ‘Experience’ an exotic notion
2. Not apt for evidence
  - (a) *Overflow* (Ryle 1949, 157; Rosenthal 1986; Block 2011)—questions about experience remain unresolved
  - (b) *Modes of presentation* (speckled hen; ‘I have the experience so-and-so had when such-and-such’; dualism: Stalnaker 2008, 100–1)—some questions unresolved, some incorrectly resolved
  - (c) *Misfire* (Chalmers 2010b, 269–70; Hawthorne 2006 v Chalmers 2010c, 173)—some questions incorrectly resolved

#### 3.2 Mode evidentialism

⇒ One has evidence whether  $\varphi$  just if one’s belief whether  $\varphi$  is held under the ‘evidential mode of presentation’

*def* A belief with content  $p$  is held under the *evidential mode of presentation* just if one’s holding it consists in one’s occupying the  $p$ -cell of a sensible-quality partition imposed by attention

*def* *Sensible qualities* include, perhaps:

- (i) Conditions of pressure and temperature in one’s body, along with its itches and tingles and other ‘sensations’, and its intrinsic shape; perhaps, the conditions of receptors involved in sight and hearing (‘organismal’ qualities)
- (ii) Colors, sounds, shapes, and textures (‘environmental’ qualities)
- (iii) Positions and distances relative to one, ‘wide’ conditions of sensory stimulation like *seeing this and that (with double vision)* or *hearing thus and so (through stuffed ears)* (‘ecological’ qualities)

##### Issues

1. Nothing exotic
2. Apt for evidence
  - (a) *Overflow*—subject-matter *refines* mode
  - (b) *Modes of presentation*—built in
  - (c) *Misfire*—evidence versus interpretation

#### 3.3 Mode evidentialism and externalism

- The subject-matter approach is traditionally associated with *internalism*: after all, we can be uncertain or even wrong about pretty much anything at all familiar, so the tradition was required to exoticize the ‘revealed’ subject-matter; and the greater load of theoretical perplexity stemming from psychology, relative to from the ‘external world’, then made it convenient to locate this exotic subject-matter in the ‘internal’

- But if we ditch the subject-matter approach for the mode approach, uncertainty and error are dealt with *structurally*: there is no longer any need to impose any restrictions whatever on the subject-matter of evidence
- One could, if one liked, restore an ‘internalist’ flavor by excluding the environmental and ecological from ‘sensible qualities’, thereby restricting to the organismal; but organismal qualities are still not ‘internal to consciousness’, so the notion of ‘internality’ is off target; and this would draw a line unrecognized by common sense and unmotivated by theory—so best to avoid it

## 4 Expressivism, purport-qua, fragments

A quick sketch of mental-discourse expressivism:

### 1. Faultless disagreement in endorsement-logic

- (a) Rather than saying a proposition  $p$  is *correct* (*incorrect*) just if  $p$  is true (false) at the actual world, we say a mental condition  $c$  *treats*  $p$  as *correct* (*incorrect*) just if  $p$  is endorsed (antiendorsed) by  $c$ —namely, if the belief-content of  $c$  entails  $p$  (anti-entails  $p$ —entails not- $p$ )
- (b) A mental condition  $c$  endorses a sentence  $\varphi$  just if  $c$  endorses  $\llbracket\varphi\rrbracket^c$ , the  $c$ -content of  $\varphi$
- (c) When  $c$  endorses  $\varphi$  because  $c$  has feature  $F$  and  $c'$  antiendorses  $\varphi$  because  $c'$  has feature  $G$ , this is *disagreement* just if no mental condition can have both  $F$  and  $G$
- (d) When  $c$  and  $c'$  disagree over  $\varphi$ , this is *faultless* just if  $c$  treats  $\llbracket\varphi\rrbracket^{c'}$  as incorrect and  $c'$  treats  $\llbracket\varphi\rrbracket^c$  as correct

### 2. Simulation, purport-qua, and mental discourse

- (a) Simulation as a Stalnaker conditional: my beliefs about Fred’s mental condition = my ‘implicit simulations’ of Fred = the condition I *purport-qua-Fred* to inhabit = an implicit suppositional

mental condition arrived at by minimally mutating my own condition to that condition with the exact evidence-content I treat Fred as having (supervalue when I am uncertain about Fred’s evidence-content) = the Fred-adequate mental condition most available from my own

- (b) An ‘availability-distance’ metric  $\alpha$  relating mental conditions, where low values of  $\alpha(c, c^*)$  go with greater ease from the point of view of  $c$  in empathizing with  $c^*$ ;  $\alpha(c, c) = 0$  (with insurmountable difficulty,  $\alpha(c, c^*) = \infty$ )
- (c) Let  $c$  treat  $j$ -at- $t$  as having (exactly) the evidence-content  $e^*$ : then  $c^*$  is  $c$ -adequate for  $j$ -at- $t$  just if  $c^*$  has  $e^*$
- (d) Stalnaker’s Assumption: some unique  $c$ -adequate for  $j$ -at- $t$  condition is at a least distance from  $c$  (or else every  $c$ -adequate condition is at infinite distance)
- (e)  $c$  purports qua  $j$ -at- $t$  to  $\Psi$  (not to  $\Psi$ ) just if, for that least (finitely) distant  $c^*$  from  $c$  such that  $c^*$  is  $c$ -adequate for  $j$ -at- $t$ ,  $c^*$   $\Psi$ s (does not  $\Psi$ )

### 3. Faultless disagreement in purport-qua

- (a) The  $c$ -content of ‘ $j$ -at- $t$   $\Psi$ s’ tests  $c$  for purport-qua  $j$ -at- $t$  to  $\Psi$ : the trivial proposition just if  $c$  does so purport, and the impossible proposition otherwise
- (b) Let  $c \neq c'$  and agree that Fred-at- $t$  has evidence  $e^*$ ; and let  $c^*$  and  $c^{**}$  be  $c$ - and  $c'$ -adequate, and also the least  $c$ -distant and least  $c'$ -distant such conditions: then still, perhaps  $c^* \neq c^{**}$   
(Intuitively, if each of you and I try to imagine our way into Fred on the basis of the same beliefs about Fred’s evidence, the difference in our starting points might lead to differing endpoints)
- (c) If so, then for some  $\Psi$ ,  $c$  purports qua Fred-at- $t$  to  $\Psi$  but  $c'$  purports qua Fred-at- $t$  not to  $\Psi$ —so  $c$  endorses ‘Fred-at- $t$   $\Psi$ s’ while  $c'$  antiendorses it
- (d) That is *disagreement*, because it is impossible to purport qua Fred-at- $t$  both to  $\Psi$  and not to  $\Psi$ ; but it is *faultless*, because  $c'$  treats the  $c$ -content of ‘Fred-at- $t$   $\Psi$ s’ as correct and  $c$  treats its  $c'$ -content as incorrect

#### 4. Fragmentation and quasi-purport

- (a) When no  $c$ -adequate for Fred-at- $t$  condition is at a finite  $c$ -distance, Fred-at- $t$  is not perfectly intelligible for  $c$ : instead, he is only imperfectly intelligible in a range of conflicting ways—the fragments; the idea is that the fragments arise when we attempt to ramify evidence through strategies of interpretation but for whatever reason find we cannot encompass all of it: in that case, we drop out some evidence here, some there, and adopt multiple conflicting interpretations
- (b) Because mode-evidence is always the true cell of some partition, any total evidence-content  $e^*$  is a ‘granulated proposition’ (Helie 2016): a partition  $\pi(e^*)$  paired with a proposition  $p(e^*)$  discriminated by  $\pi(e^*)$  (the former is the union set of some subset of the latter)
- (c) Let  $e^\dagger$  be  $e^*$ -admissible just if (i)  $\pi(e^*)$  refines  $\pi(e^\dagger)$ ; and (ii)  $p(e^\dagger)$  is the strongest proposition discriminated by  $\pi(e^\dagger)$  and entailed by  $p(e^*)$
- (d) When  $e^\dagger$  and  $e^\ddagger$  are both  $e^*$ -admissible, let  $e^\dagger$  be as  $e^*$ -nice as  $e^\ddagger$  just if  $\pi(e^\dagger)$  refines  $\pi(e^\ddagger)$
- (e) To reduce clutter in the following definition, assume when relativizations are not explicitly stated that we are discussing  $c$ , Fred-at- $t$ , and the evidence  $e^*$   $c$  treats him as having:  
Let  $C^{fr}$  be the set of fragments (in the view of  $c$ , for Fred-at- $t$ ):  $C^{fr}$  contains  $c^\dagger$  just if (i)  $c^\dagger$  is at a finite distance; (ii) no finite-distance condition has nicer evidence; (iii) no condition with the same evidence is less distant
- (f) Where  $c^\dagger \in C^{fr}$ ,  $c^\dagger$ -quasi-purports qua Fred-at- $t$  to  $\Psi$  just if  $c^\dagger \Psi$ s;  $c$  purports qua Fred-at- $t$  simpliciter to  $\Psi$  just if for every  $c^\dagger \in C^{fr}$ ,  $c^\dagger \Psi$ s (exceptionless quasi-purport); and  $c$  treats Fred-at- $t$  as incoherent in regard to  $\Psi$  just if  $c$  purports simpliciter neither to  $\Psi$  nor not to  $\Psi$

#### 5. Fragments in practice

- (a) Thinking of Fred as fragmented is a kind of *ineliminable uncertainty* about Fred’s mental state: fragments arise when I have

*more information* about Fred than I can make sense of; getting still more can only make things worse (feeding purport with *exact* evidence-contents highlights this)

- (b) But imperfect intelligibility is also *better* in a certain respect than uncertainty: each fragment has its own internal coherence, and is lashed to a certain bundle of evidence; accordingly, I can use these fragments adventitiously, holding each of them up to Fred ‘in the moment’ to see which is then apt as an explanatory template
- (c) Accordingly, I may for certain purposes reason about fragment  $c^\dagger$ , then later for distinct purposes reason about fragment  $c^\ddagger$ , and wind up disagreeing with myself between the course of  $c^\dagger$ -purport and the course of  $c^\ddagger$ -purport
- (d) This is not a kind of ‘many-subjects’ view, however, because I only am able to deal with fragments individually when I mutilate some of my information: the fragments are essentially creatures of pretense, and do not survive my commitment to literalism; but the problem with my doing *that* is that what it leaves me with is not very useful.

## References

- Block, Ned, 2011. 'Perceptual consciousness overflows cognitive access'. *Trends in Cognitive Sciences*, 15:567–575.
- Chalmers, David J., 2003. 'The content and epistemology of phenomenal belief'. In Quentin Smith and Alexander Jokic, editors, *Consciousness: New Philosophical Essays*. Oxford: Oxford University Press.
- Chalmers, David J., 2010a. *The Character of Consciousness*. New York: Oxford University Press.
- Chalmers, David J., 2010b. 'The content of phenomenal concepts'. In [Chalmers 2010a](#), chapter 8. Oxford: Oxford University Press. Abridged and updated from [Chalmers 2003](#).
- Chalmers, David J., 2010c. 'The two-dimensional argument against materialism'. In [Chalmers 2010a](#), chapter 6. Oxford: Oxford University Press.
- Collingwood, Robin George, 1946/1993. *The Idea of History*. Oxford: Oxford University Press.
- Elga, Adam and Agustín Rayo, 2009. 'Fragmented belief states and logical omniscience'. Talk at CUNY Graduate Center.
- Hawthorne, John, 2006. 'Direct reference and dancing qualia'. In Torin Alter and Sven Walter, editors, *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.
- Heal, Jane, 2003. *Mind, Reason, and Imagination*. Cambridge: Cambridge University Press.
- Hellie, Benj, 2007. 'Higher-order intentionalism and higher-order acquaintance'. *Philosophical Studies*, 134:289–324.
- Hellie, Benj, 2011. 'There it is'. *Philosophical Issues*, 21:110–164.
- Hellie, Benj, 2013. 'It's still there!' In Richard Brown, editor, *Consciousness Inside and Out*. Dordrecht: Springer.
- Hellie, Benj, 2016. 'Rationalization and the Ross Paradox'. In Nate Charlow and Matthew Chrisman, editors, *Deontic Modals*. Oxford: Oxford University Press.
- Hellie, Benj, 2017. 'An analytic-hermeneutic history of Consciousness'. In Kelly Michael Becker and Iain Thompson, editors, *Companion to the History of Philosophy, 1945–2016*. Cambridge: Cambridge University Press.
- Horgan, Terence and John Tienson, 2002. 'The intentionality of phenomenology and the phenomenology of intentionality'. In David J. Chalmers, editor, *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press.
- Lewis, David, 1972/1997. 'Why conditionalize?' In [Lewis 1999](#). Handout from 1972 with a 1997 'Introduction', first published in this collection.
- Lewis, David, 1982. 'Logic for equivocators'. *Noûs*, 16:431–441. Reprinted in [Lewis 1998](#).
- Lewis, David, 1996. 'Elusive knowledge'. *Australasian Journal of Philosophy*, 74:549–67. Reprinted in [Lewis 1999](#).
- Lewis, David, 1998. *Papers in Philosophical Logic*. Cambridge: Cambridge University Press.
- Lewis, David, 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.
- Rayo, Agustín, 2013. *The Construction of Logical Space*. Oxford: Oxford University Press.
- Rosenthal, David M., 1986. 'Two concepts of consciousness'. *Philosophical Studies*, 49:329–59.
- Ryle, Gilbert, 1949. *The Concept of Mind*. London: Hutchinson and Co.
- Stalnaker, Robert C., 2008. *Our Knowledge of the Internal World*. Oxford: Oxford University Press.