Approaches to Intentionality
By William Lyons

In this clearly written, informative book Lyons provides a critical survey of some contemporary realist theories of intentionality (Part I) and also sketches his own theory—a layered, developmental 'in-the-head' realism (Part II). Lyons adopts a both/and approach: intentionality is in the brain, in consciousness, and also a feature of language, though different forms of intentionality are at work in each case. Intentionality is not just a linguistic phenomenon, nor is it purely a brain matter; Lyons argues for gradated levels: brain, pre-linguistic consciousness, and adult language-user, with fully fledged propositional attitude intentionality only operative at the highest level.

Part I is a reliable, up-to-date, critical survey of five leading contemporary accounts of intentionality, instrumentalism, representationalism, teleology, information processing and functionalism, exemplified by Dennett, Fodor, Millikan (coupled with McGinn), Dretske and Loar. Lyons begins by situating Dennett's instrumentalist approach historically in the tradition of Comte, Carnap and Quine. According to this tradition, there are only physical transactions in the world, so talk of intentionality is 'just talk' (p. 38), extremely useful, perhaps even indispensable for humans, but not ultimately part of the furniture of the world. As Lyons sums up this view:

"The mind has no intentionality for the simple reason that there is no mind, and the brain has no intentionality because it has neither states nor processes with content. Intentionality is merely a feature of a particular part of our language plus a strategy which generates that 'language game'."

(p. 27)

Lyons sees Dennett's instrumentalism as in one sense 'reductionist' and, in another sense, simply 'dismissive' (p. 36) of intentionality. Lyons argues that Dennett is required to interpret the success of the intentional stance as really a lucky guess, whereas a more reasonable position would recognize that our predictive successes must be based on real knowledge inside the head of the interpreter. Instrumentalism, for Lyons, has done away with in-the-head content for no good reason.

Representationalism (Fodor), on the other hand, is an 'industrial strength' (p. 55) 'in-the-head' realism about intentionality and mental content. Lyons presents Fodor as a rationalist and nativist heavily influenced by Chomsky. Lyons agrees with Fodor's view that the brain is not just a syntactic but a real semantic engine, driven by intentional states. However, Lyons parts company with Fodor on the postulation of a language of thought (LOT). For Fodor propositional attitudes are in the brain itself; the brain possesses a 'stiffened up version of our folk psychology' (p. 55). Lyons criticizes Fodor for being too literalist about LOT and here sides with the instrumentalist view that folk-psychology is a cultural artefact. However, he disagrees with the instrumentalist claim that there is no analogue for propositional attitudes and contents in the brain. There is some basis in the brain for Lyons, but not as Fodor envisions it. Lyons criticizes Fodor not just for his nativism but also for having no account of consciousness or of the learning process. Furthermore, as Lyons will argue in Part II, there are non-representational, information-bearing states, e.g. 'grasps of and contacts with' (p. 62).
Lyons is clearly attracted to some biological basis to intentionality. In his discussion of Millikan's teleological account, he agrees with her that intentionality is 'an objective natural feature of humans' (p. 76), grounded in external natural relations, and understood in terms of the biological notion of proper function. Proper function is what the organ is designed by evolution to do, as it operates under normal conditions in a standard environment. So, Millikan holds, in Lyons' terms, 'believing is the activity of a device which is designed by evolution to have the effect of producing true beliefs in the believer' (p. 77). The content of a belief is not in the head but rather is deciphered by what someone possessing the belief is supposed to do. The intentionality of a belief cannot be read off here and now from any physical incarnation of the belief itself; beliefs are what they are because of what they are supposed to do, as conditioned by their evolutionary history. Function fixes content, as McGinn puts it, who, as Lyons shows, offers a version of Millikan's teleological-cum-biological account. Nevertheless, for Millikan, the beliefs themselves are in the head, and she agrees with Fodor that our folk psychology does carve nature at its joints (p. 81). Lyons' main difficulty with Millikan is akin to Fodor's objection (in the Donnellan lectures given at Trinity College, Dublin in 1989), namely, that evolution is too crude an instrument for individuating mental content - no more than an act of faith, as Fodor says. How could evolution give us a fine-grained account of our present beliefs? Millikan's approach gives everything with evolutionally proper function (e.g. heart or lungs) a content in just the same way as beliefs. Ultimately this appeal to evolution is too generalized to be informative.

In his chapter on Dretske's information-processing approach, Lyons praises Dretske's account of content as 'sophisticated and plausible' (p. 97) yet concludes that it does not do justice to the epistemic and conscious aspects of our intentional life. Dretske approaches content in terms of the amount of information that gets communicated. Any situation with possibilities is a possible source of information. For Dretske it is 'information that manufactures meaning, not behaviour' (p. 105). Information is the beginnings of intentionality, for Dretske; indeed, it is original intentionality. Dretske distinguishes different kinds of representational systems, but, for him, humans are genuine representers with genuine content. Lyons thinks Dretske too closely identifies neuronal structures with representings, and he is critical of Dretske's need to postulate self-understanding internal representations. For Lyons, though the brain does have a primitive intentionality of its own, whatever is in the brain is not the kind of thing we should call a 'belief' or 'desire' (p. 115). The information processing is done by the brain without ever knowing it as information, in Lyons' view. Lyons further holds that Dretske's account cannot adequately handle the issue of consciousness.

The chapter on Brian Loar, presenting him as a pure functionalist, is welcome. For Loar, intentional states are real physical states with real causal powers. There is type–type identity between propositional attitudes and brain states, though there is enormous plasticity in the manner in which attitudes are realized physically. Loar needs no LOT and no representations in the brain as such. The brain has a network of states with causal interrelations isomorphic with the logically interrelated network of propositional attitudes. Having meanings in the head is having physical states with non-intentional causal roles. Lyons is sympathetic to the view that the physical states are not precise analogues of propositional attitudes, and also to the functional role account of the attitudes, but he is sceptical about the brain level echoing 'the logical-rational network of propositional attitudes as defined by the abstract functional role theory' (p. 142). Lyons has a deep feeling that we won't find brain equivalents of our ordinary attitudes, and this is because these attitudes are constituted socially, as he develops in Part II.
In Part II Lyons sketches his gradualist, layered view of intentionality. Aboutness comes in layers, though Lyons' stages of life's way are philosophical and not actually developmental. At the most primitive level there is brain intentionality. The brain is the repository of analogue records which, contra Fodor, should be explained without recourse to LOT (p. 168). As Lyons puts it (p. 169), the brain reduces all stimuli into a common currency but not into a common language. Lyons distinguishes between a language, a code and an 'analogue transmission device' (p. 164) where the latter is 'a vehicle for transforming in a causally covariant analogue manner, certain properties of one physical device into another physical device'. Though Lyons does not invoke the term, this is what engineers call a transducer, something which transforms but does not translate (e.g. a telephone). The network merely stores a record, e.g. a pattern of light; nothing in the brain represents the connection between this pattern and the environment, though there is plainly a causal connection. In Lyons' example (p. 168): if my nose is punched, my nose stores a record of some aspect of the punch, but does not represent the punch (Lyons defines representation as 'artefacts made to resemble what they stand for according to some convention' (p. 175)). Lyons wants clearly to distinguish a record from a representation, whereby representation has some kind of resemblance, albeit conventionally acquired. But, since Descartes, modern philosophy has denied that representation requires resemblance and has gone instead for a semiotic or symbolic account of representation – it is not the object itself which matters, but the lawlike manner it stands in relation to the other object. Lyons distinguishes the causal co-variance of records from the symbolic representation of symbols. Brain information is primitive or raw information which causes responses. Representations and symbols need an interpreter while records don’t and Lyons denies, citing Edelman in support, that there are any grounds for believing that the brain can be an interpreter of its own analogue records. There is no interpretation at the level of brain chemistry, but surely there is a lot more to be said about the emergence of symbol interpretation.

The brain processes uninterpreted contents, but contents nonetheless, for Lyons, who defines content simply as something contained in the brain which is causally co-variant with the stimulus and guides action. Records function in 'an unmediated causal way', (p. 172). However, attributing uninterpreted content to the brain is problematic: surely the very notion of content suggests aspect or point of view, and hence is already interpreted. Uninterpreted contents wouldn't be contents at all – just goings on. Though the brain is a 'true semantic engine' for Lyons, it is not of itself an epistemic engine (p. 157). This level of signification is arrived at only on the purely human plane of acculturation and language. Moving meanings about the brain is processing them but not interpreting them; interpretation requires the human world of language. Lyons wants this uninterpreted content to be the basis of higher-level interpretation which takes place at the human level, but the problem is precisely to explain how uninterpreted brain transactions turn into semantic entities. Lyons' account does not provide a satisfactory account of how the postulated layers are linked. What are the bridging factors?

For Lyons, the story gets progressively more complex when we add in the notion of sensory experience. Here he introduces his account of infant intentionality. Memory plays a key role in giving sensory experiences a certain orientation. For Lyons, the infant has a behavioural version of intentionality; one stimulus inevitably is followed by another so that one is, in some sense, 'about' the other (p. 162). The infant lives in a kind of Proustian reverie of associations, but given memory, the infant is slowly able to make associations among repeated happenings, and begin to adjust her responses accordingly. The infant has perceptual discriminations, even
‘proto-concepts’, before she has concepts. The infant’s sensory experiences are intentional in that they are discriminatory. There is a kind of proto-conceptual intentionality in the infant’s awareness interpretation – characterized as ‘Brentano without the Aristotelian-Cartesian underpinning’ (p. 159). But what precisely accounts for the aboutness of the experience? Ability to discriminate surely is an indication of the aboutness, not an explanation of it. Lyons’ answer is that the aboutness is built up in a Humean manner by repeated associations stored in memory. But again, this does not tell me how it happens: it just rephrases the way it does happen in language a Humean is happy with. Don’t we need more in the way of an account of the child’s attitudes, her intentional acts? Lyons is surely right that the infant cannot be said to have propositional attitudes, as her attitudes cannot have propositional form. She has, rather, non-propositional contents (p. 159). But Lyons claims that, when language users ascribe propositional attitudes to themselves and others, they are in fact picking out ‘real features of real humans’ (p. 163). What are the real features being picked out in infants corresponding to the attribution of attitudes, albeit non-propositional ones? Infants learn that one thing can have significance through being ‘constantly conjoined’ with another thing according to Lyons (p. 158). But how? How does a child know that her sensations are about the world? Here Lyons dismisses both nativist and standard empiricist accounts (incidentally, renouncing his own earlier empiricism). We cannot assume that something like Humean vividness of impressions indicates their extra-mental origin, even combined, as it is in some modern versions, with phenomenal awareness of the causal link itself. So empiricism won’t do. Lyons’ answer is that all we need is that the child respond in an appropriate way; we don’t need to attribute concepts to the child. The responses are honed by evolution and by developmental progress as mapped by developmental psychology (Lyons’ account here is like Dennett’s ‘ABC’ learning in *Kinds of Minds*). Though I agree with Lyons’ criticism of empiricist accounts of the development of early human intentionality, I feel that Lyons underplays the strength of nativist accounts. His description of the child’s development avoids attributing built-in conceptual apparatus to the child, but much of his account seems so close to behaviourism that I cannot see how it avoids behaviourism’s well-known pitfalls. He also seems more intent on classifying the kind of content in the infant’s mind/brain than the kind of intentional attitudes with which she operates. But a full story of intentionality must cover both acts and contents.

In the course of his account Lyons also tentatively ventures an approach to consciousness which in fact sounds quite Searlian, though he does not invoke Searle and departs from Searle in holding on to mental causation. He accepts that consciousness is a natural product of evolution, and hence can be naturalized, but stresses that consciousness also belongs to our folk-psychology, and that its contents are culturally constructed through language using. Lyons accepts a certain token-token identity between conscious events and brain events, but stresses that each of us will have different tokenings because of the different ways our brain networks have developed and been enforced by environment. Lyons sees consciousness as a product of brain processing rather than a segment of the brain (p. 185). He wants to preserve consciousness as knowable in a first-person way, and entertains, but ultimately rejects, subjective and objective access to the same brain analogue record, what he calls the ‘duality of access’ account. Lyons assumes that these records at a high level have an internal subjective point of view (but does not this go against his earlier insistence on the uninterpreted status of brain transactions in themselves?). He is, in short, accepting some form of emergence, but does not offer any detailed account. Here engagement with Searle would have helped the reader to situate Lyons’ account.
In Lyons' account, full intentionality requires interpretation, and hence true epistemic intentionality is a feature of our social world, of our intersubjectivity, requiring high-level language- and symbol-using ability. At this fully developed level there is independence of content and object, with propositional attitudes having perspectival intentionality. Humans invent their own representations; they don't find their brains lumbered with them. Nevertheless, intentional description picks out real features in us (p. 163) – this starts to sound like Dennett's pragmatic realism of true believers (though my own view is that the term 'real' is being stretched too far by Dennett, whereas it is left fairly undefined by Lyons). But how can this process get started? Don't we need more than analogue records in the brain to explain the intrinsic intentionality which is required to support linguistic intentionality? I don't find that Lyons' account (in Chapter 7) of the emergence of the interpretative level from the pre-intentional informational level of the brain gives me enough material.

Part I is an excellent, informed review of some of the strongest contemporary theories of intentionality. Lyons' easy prose style conveys a deep knowledge of the area in a very palatable form, together with solid, common-sense criticisms of all the candidate theories. Part I is, inevitably, more successful than the more suggestive second part, which attempts a coherent theory of intentionality. The book's strength is its clear, accessible, yet thoughtfully critical, portraits not just of the more familiar theories of Dennett and Fodor but also of the less widely known work of Millikan and Loar. Lyons' choice of intentional theories is unusual and distinctive (he deliberately omits Searle, Putnam, Davidson and Chisholm, for example), but Lyons, by comparing and contrasting the chosen current theories, provides an accurate and thoughtful guide to a fairly uncharted yet very important domain of contemporary philosophy of mind.

Lyons' own theory in Part II is more difficult to situate in the line-out of the usual suspects on intentionality. Though flagged as a 'different approach', it is not so much different as a careful selection of what he takes to be best in recent accounts. Lyons, like Searle, insists that too many authors have sought to treat intentionality as a problem which must be resolved or preferably dissolved rather than a genuine fact of the world which has to be explained. Lyons does not agree with the view which sees intentionality as a purely linguistic or grammatical phenomenon (p. 163). He argues for the view that the brain alone has a very weak intentional power – processing without interpretation. In agreement with Millikan, Lyons is in no doubt that evolution has given rise to the brain's intentional powers, but these powers are weak until taken up and reinterpreted in the human world. Lyons insists, with Fodor (and also Putnam), that evolution cannot explain our fully fledged intentional usage. With Fodor and Dretske, he holds that the brain is a real information processor, a true semantic engine. Against Dretske, Lyons takes the view that representations don't understand themselves. With Searle (and also Putnam) he emphasizes the socially constituted aspect of full human intentionality.

In the end Lyons tries to have everything – evolution, content-bearing states 'in the head', intentionality as a natural relation between brain and world, even a relaxed behaviourist account of concept acquisition at the infant stage. The great merit of book is its easy, no-nonsense conversational style and its crystal-clear expositions of difficult thinkers like Loar and Millikan, as well as its emphasis on the need to have a multi-dimensional account of intentionality. There is a certain amount of repetition of key ideas from chapter to chapter which is actually helpful for the reader. The chief weakness is Lyons' reluctance to give us a metaphysical account of attitude possession and of mental acts themselves, and
perhaps an over-optimistic assumption that upper-level intentionality can be accounted for horizontally in terms of a socio-linguistic account and vertically in terms of brain functioning, without worrying about the complexities of horizontal and vertical causation. All in all, though, an excellent critical guide to a key area of contemporary philosophizing which will considerably benefit not just novices in the area but the professionals also.

University College Dublin

Benacerraf and his Critics
Edited by Adam Morton and Stephen P. Stich

Paul Benacerraf is an unusual phenomenon in the contemporary philosophical world. Despite his having spent a number of decades at the forefront of his discipline, his published work amounts to a handful of papers. Each paper has, however, been a major contribution to its field, and two - 'What Numbers Could Not Be' (WNCNB) and 'Mathematical Truth' (MT) - have become landmarks in the philosophy of mathematics: very few writers have felt it safe to ignore the challenges articulated there. In fact, Benacerraf's thought is more readily identified in terms of a set of challenges and an underlying methodology than in terms of substantive philosophical views. True, WNCNB does include an espousal of a kind of structuralism about mathematical objects, but even here Benacerraf admits that this was a transient feature of his philosophy and he never gave the position the sort of systematic development devoted to it by others (e.g. Hellman, Parsons and Resnick). This perhaps accounts for the manner in which this book diverges from the series' form. Books in the series generally contain a number of essays on the chosen philosopher followed by his/her responses. Here the collection begins with an essay by Benacerraf himself (he becomes his own most acute critic), followed by a collection of essays, some of which touch only tangentially on Benacerraf's work (one - Steiner's - is on Wittgenstein, not Benacerraf) and some with a moral which is supposed to be, in some manner, Benacerrafian. Although many of the essays are very good, the sum total is a book which lacks overall coherence. In part this may have been a consequence of striving to include only new material. Benacerraf's work has been so much discussed that contributors would have been hard pushed to comment further on central aspects of that work. (For lack of space I've been forced to be selective about which chapters to discuss.)

Benacerraf's chapter divides into three. It begins with an engaging depiction of the philosophical climate out of which his two seminal papers grew. He vividly portrays the scientifically inclined, positivistically influenced philosophy which dominated America and contrasts this with that influenced by Wittgenstein and the ordinary-language philosophers in Britain. What emerges is an interesting account of the differing views about the nature of philosophy, the place of research programmes and analysis. A crucial factor in determining these differences is alternative conceptions of the a priori. Having provided us with this much context of the perceived role of reductive programmes, Benacerraf goes on to offer a critique of his WNCNB. His conclusion, thirty-something years on, is more tentative. The argument might be responded to in any of three ways: (i) in something of the fashion recently investigated by Wright, Hale and Boolos; (ii) by a resolute realist who insists that despite our inability to arbiter between competing reductions one