JOHN DILWORTH

THE DOUBLE CONTENT OF PERCEPTION

ABSTRACT. Clearly we can perceive both objects, and various aspects or appearances of those objects. But how should that complexity of perceptual content be explained or analyzed? I argue that perceptual representations normally have a double or two level nested structure of content, so as to adequately incorporate information both about contextual aspects Y(X) of an object X, and about the object X itself. On this double content (DC) view, perceptual processing starts with aspectual data $Y'(X')$ as a higher level of content, which data does not itself provide lower level X-related content, but only an aspectually encoded form of such data. Hence the relevant perceptual data $Y'(X')$ must be 'de-contextualized' or decoded to arrive at the X-related content X', resulting in a double content structure for perceptual data, that persists in higher-order conscious perceptual content. Some implications and applications of this DC view are also discussed.

I have a very basic thesis to propose, namely that the content of perception is not ordinarily confined to a single level or kind of content, but instead that it usually involves a *double* or *two* level content structure – a *double content* (DC) view of perception. I shall also propose a schematic account of the kind of basic cognitive processing that could implement the relevant representational perceptual structures.

To begin, here is a kind of 'complexity of the stimulus'[1] argument for a two level content view of much perceptual representation. One of the functional requirements on a concept of perceptual content is that it should adequately account for our actual abilities to perceptually identify and re-identify objects (including events, states of affairs and so on) in our perceptual fields. For example, it is possible for a person P to perceive an object X as being the same object even if it is rotated, moved closer or further away, presented under different lighting, and so on, or in general to recognize different contextual *aspects* of an object X as being aspects of the *same* object X.[2]

Thus person P's X-related perceptual content X' must in some way remain *the same* through these changes in her perceptual environment – at least in a functional sense of triggering or maintaining the same recognition or identification processes that result in the object continuing to be identified as being object X.

However, at the same time P's perceptual contents must also adequately reflect the relevant aspectual changes as object X is rotated, moved, differ-

Synthese (2005) 146: 225–243
DOI 10.1007/s11229-004-6209-3

ently lighted and so on. Thus P must be able to perceive different *aspects* A1, A2 . . . An of object X, while nevertheless continuing to perceive them as aspects of *the same* object X.

I claim that single-level accounts of perceptual content – no matter how they are analyzed with respect to the contentious issues concerning wide versus narrow content – are not able to adequately account for this complex structure of re-identifiable perceptual contents, hence showing the need for a two level content view. Person P's perceptual content with respect to object X is neither simply object X itself – which would mean that the different perceptual aspects of X were not adequately perceived by her – nor simply the different *aspects* of X, which would result in a lack of perception of them as all being aspects of the *same* object X. Instead, in some way P must perceive *both* the aspects, *and* the fact that they are aspects of the same object X, and hence her perceptual contents must in some way include *both* components during her perception of any particular aspect of X.

As to why a two level content view is needed, rather than simply a more complex single-level view, it would not help to simply view perceptual contents as including both an aspect-shorn X-related content, plus an X-shorn aspectual content: clearly an adequate theory must in some way view aspectual content as itself being about X. I shall show how a two level content view could do this.

My proposal is to do so by making use of the logic of representation: perception of X involves not simply a perceptual representation of X, but instead a perceptual representation whose content includes both information about *aspects* of X, and about X itself, in a *two level*, nested content structure.

1. THE LOGIC OF SECOND ORDER REPRESENTATIONS

In order to clarify my proposal, it will be useful to start with a consideration of *second order* representations – representations *of* representations – such as a perceptual representation A (which is some concrete perceptual state of a person) of a concrete picture B, that in turn represents some object X (which picture B is hence itself a representation), or another concrete picture C which is a picture of a further picture D that represents some object Y.[3]

As a preliminary, I shall make the commonplace assumption that if a representation A represents some item B, then A has a *representational content* B′ consisting of, or associated with, the properties that A represents B as having, so that B′ is the B-related content of A. To begin, consider

the first case mentioned above, namely that of a second order perceptual representation A of a picture B that represents an object X. In order to adequately represent both the picture B and the fact that it represents object X, presumably the content of A must include both its B-related content B′, and the X-related content X′. Question: what is the relation of the two contents B′ and X′?

Now B′ is the B-related content of A, and X′ is the X-related content of A. But it would be wrong to think of contents B′ and X′ as being totally independent contents, since in normal perception of a picture B, the picture-related perceptual content B′ – i.e., the properties that one's perceptual state A represents picture B as having – must include not only its physical properties as a pictorial artifact, but also picture B′s own X-related content properties X′ – for otherwise the content B′ related to picture B would fail to be genuinely *pictorial* content. Thus in some way content B′ must include or subsume content X′.

However, content X′ itself consists only of the properties that picture B represents object X as having, so that it does not itself include or subsume content B′. Hence there is an *asymmetric* relation between contents B′ and X′: B′ subsumes X′, but not vice-versa. I shall describe the resulting relation between contents B′ and X′ as being one of two *levels* of content, in which content B′ is at a *higher level* than content X′, and in which the *lower-level* content X′ is *nested* inside the higher-level content B′.

To be sure, in the absence of a good, well-articulated theory of the nature of content, and of representation generally, the precise relations between contents B′ and X′ remain unclear,[4] but for present purposes the important point is that there is an *asymmetry* between B′ and X′, which asymmetry may usefully be expressed in terms of the current two level, nesting terminology.

The next issue about second order representations concerns the precise meaning of the claim that they are representations *of* representations. That claim might seem to imply that there are *two* distinct representational entities in such cases: that if object A represents B, which in turn represents C, then object A does the higher level representing, while object B does the lower level representing.

However, such an analysis would ignore the fact that second order representations involve only *one* concrete object or state: thus, in the initial perceptual case where perceptual state A represents picture B that in turn represents object X, it is perceptual state A alone that has to represent *both* picture B *and* its represented object X, since the actual physical picture B that represents X is not itself an ingredient in the relevant second order

representation, which instead only involves the B-related *content* B′ plus its nested content X′.

Thus strictly speaking there cannot be any second or higher order representations that involve actual or literal distinct stages or levels of representation. To be sure, there can be *chains* of representations, in which for example perceptual state A represents concrete picture B, which concrete picture in turn represents object X, with genuine representation involved at each stage; but such a two-item representational chain is not itself a single second order representation.

Thus I propose instead the following straightforward account of a second order representation in which a perceptual state A represents a picture B that itself represents an object X. In such a case, state A represents both picture B and object X, and A's content consists of a higher-level content B′ and a lower-level nested content X′. Thus on this account, cases of higher order representation are cases of multiple representation of entities, one for each level of nested content.

2. WIDER USES FOR NESTED CONTENT

Now that some basic elements of a theory of second (and higher) order representation are in place, the results can be generalized so as to apply to a wider range of cases. In particular, I shall argue that the nested double content analysis of second order representations may be applied also to cases of perception of aspects and objects that do *not* stand in a relation of representation to each other.

As a preliminary, I shall accept the common assumption that representational perceptual states are the causal result of incoming, sub-personal or unconscious perceptual data items received through the senses, which sub-personal data items are closely associated with the relevant personal or conscious perceptual contents of such perceptual states.

Recall from the Introduction to this paper that a person P, in perceiving an object X, must both be able to perceive the same object X, and also the various different aspects or contextual changes in the appearance of X, as the viewing conditions for object X are varied. Thus P must be able to perceive different *aspects* A1, A2 . . . An of an object X, while nevertheless continuing to perceive them as aspects of the *same* object X.

Now in general, aspects of an object X, or contextual conditions under which an object X is viewed, are not themselves *representations* of X; instead, their relations to X are made up of the actual, real-world relations of different objects (including states etc.) to any given object X, along with the environmental conditions under which all of these objects interact with

each other. Nevertheless, from the point of view of a perceptual mechanism that is attempting to *represent* both the various contextually specified aspects of X, and X itself, the resulting content structure will inevitably have the same two level, nested structure as it would have if those aspects of X *did* actually represent X.

This is so because, from the stance of a perceiver who is caused to be in perceptual state S by perceptual data received in some particular situation involving object X, that perceptual state S is not directly a representation of X itself, but instead it is a representation of a *contextually modified or transformed* version of X and its features – of an *aspect* Y(X) of X rather than of X itself.

To be sure, buried within the perceptual content received will be X-related content, but my point is that the content as a whole is an *aspectual package* Y′(X′), involving intermixed content concerning contextual factors Y as well as X, which package directly, or as a whole, results in perceptual state S representing only an *aspect* Y(X) of X. Thus the content package Y′(X′) needs to be *decoded*, or be subject to an *inverse transformation*, so as to properly extract the relevant X-related content X′ from the package Y′(X′). Or in other words, the initial content Y′(X′) itself is only an *encoded* form of X-related content, rather than itself *being* X-related content.

Now if this situation is compared with a second order representation case, such as that of a perceptual state A that represents a picture B that represents an object X, if we consider just the B-related content B′ of A, and consider its relation to the X-related content X′ of picture B, there too we have a situation where the perceptual data as a whole, namely the content B′, is not itself pure X-related information, but instead it is only (informally speaking) an encoded form of the X-related information X′. That information B′ also needs to be decoded so as to extract the relevant X-related information X′.

Thus in both cases a similar content structure results, namely one in which *nesting* of content occurs: just as the content B′ of perceptual state A is a higher level content which has the lower level content X′ nested within it, so also the content Y′(X′) of a perceptual state that represents some aspect of X must have a lower level content X′ nested within it.

3. RECOGNITION AND IDENTIFICATION

The current two level, nested double content (DC) view of perception will now be reinforced by relating it to issues concerning basic kinds of functional cognitive processing of information, whose aim is to produce

perceptual *identification* and *recognition* of a given object under different circumstances.

As an initial contrast, an opposing single content view of perception would focus on *object feature matching* in its account of perceptual recognition. On such an account, a person P recognizes an object as being object X in virtue of matching the features of her perceptual content X′ with standard features of object X, such as are embodied in a stored perceptual prototype for the object. Such an account assumes that there is some finite list of necessary and sufficient conditions for some object to be X, which list is used in this matching activity.

On such an account, the changes in perceptual content that result from different aspects of an object being perceived are viewed as strictly being irrelevant to identification of the object as X, since they merely make it easier, or more difficult, to carry out the standard kind of matching process in question. For example, an object moved further away, or rotated into some unfamiliar perspective, might hence make it more difficult to identify certain features of the object, and hence more difficult to identify the object as X, but only a single level of X-related perceptual content is drawn on in any such identification or recognition cases.[5]

On the other hand, a fundamentally different, broader *contextualist* understanding of any perceptual situation is possible, which recognizes and makes use not just of X-related features, but also of *contextually relevant aspectual features Y* of the perceptual environment in which X is viewed, which features Y serve to *transform* any actual X-related perceptual data received by person P's sense organs, *prior* to her perceptual processing of that data. Or in other words, on this contextualist view, from the start perceptual processing operates not simply on X-related content X′, but on that content *as transformed by contextual factors Y*, so that the resulting perceptual content is of the form Y′(X′) – a *contextually transformed* form of X-related content X′.

What this means from both a logical and functional point of view is that a sub-personal process of feature-matching of a single level of perceptual data with a stored, stereotypical X-related list of features will no longer give an adequate account of identification or recognition of object X, because, among other considerations, the Y-related contextual transformations of the X-related raw data may have been so extreme that no match would be possible for the resulting perceptual content – or at least, not if use is made only of such a standard list of X-related features as embodied in a cognitively stored, X-related prototype.

What is needed instead is a more sophisticated matching process, which allows for and accommodates the *Y-related contextual transformations* of

X-related data that are associated with different perceptual aspects of X. So in place of a single X-related prototype for matching, instead a *series* of X-related *contextual transformation fields* should be hypothesized, each of which fields contains as its elements a related series of specialized *contextual transformations* of X-related data in some respect, so that an attempted recognitional matching of the perceptual content Y′(X′) is carried out with respect to *some element* in such a field.

For example, there could be a *spatial* transformation field, whose elements are stereotypical perceptual data that would result from some particular *spatial* transformation of object X relative to perceiver P, organized in some systematic way so that efficient searching for a match would be possible. Or another likely field might be one of *illumination* transformations, whose elements are resultant content items Y′(X′) from X-related data as transformed under different lighting conditions Y.

Clearly such X-related transformation fields would need to be integrated in appropriate ways, so that matches in each field would be partially dependent on resources from other fields, so that, for instance, a particular *spatial* transformation of object X could be matched under a range of different possible *lighting* conditions for the same object. Thus the overall process of matching the perceptual content Y′(X′) would require parallel or simultaneous matching on elements drawn from several related fields, since the contextual transformations Y of the original X-related data would presumably involve transformations in all of the relevant contextual dimensions.

A conceptually simpler model would hypothesize a single, multidimensional transformation field F1 . . . n for an object X, containing a distinct element for every possible combination of transformations Y′(X′) as applied to X-related data, such as various lighting conditions and spatial orientations, so that any perceptual content Y′(X′) could be precisely matched with some unique element in field F. Of course, the enormous storage and search requirements for such a field make it cognitively unrealistic, but the main conceptual structure associated with it would persist in more realistic cognitive approximations.

To summarize, the current contextualist view assumes that raw or subpersonal perceptual data associated with an object X, that will be processed into personal or conscious X-related perceptual content X′, does not cause a perceptual state S to provide a representation of X as such, but instead the perceptual state S that it causes provides a representation of some *contextually transformed aspect Y(X) of X*, so that the resulting higher-level perceptual content Y′(X′) is itself the content of a perceptual representation *of that contextually transformed aspect of X*, rather than of X as such.

Thus, at the sub-personal perceptual level the hypothesis is that raw, aspectually transformed X-related data cannot be immediately matched with an X-prototype, but that instead matching must proceed with elements of some appropriate X-related *transformation field*, whose elements are contextually transformed items of X-related data. When a match is achieved, the incoming perceptual data has been identified as X-related data. Then the personal level of conscious perceptual content will maintain this same representational structure, on the basis of the sub-personal contextual matching process.

4. MORE REALISTIC IDENTIFICATION METHODS

From a cognitive science rather than a philosophical point of view, the previous discussion is, at this stage, inevitably somewhat unsatisfying, in that no specific methods have yet been suggested as to how a contextual identification process, that has the relevant double content (DC) structure, might actually be carried out in an efficient enough manner so as to potentially give a plausible account of actual perceptual identification processes in humans and other animals. Here are some initial suggestions along those lines.

Recall that a conceptually simple DC model would hypothesize a single, multi-dimensional transformation field F1 ...n for an object X, containing a distinct element for every possible combination of transformations Y′(X′) as applied to X-related data, such as various lighting conditions and spatial orientations, so that any perceptual content Y′(X′) could be precisely matched with some unique element in field F. But as initially noted, of course the enormous storage and search requirements for such a field make it cognitively unrealistic, particularly since on this simple model every individual object or kind of object would presumably have to have its *own* distinctive associated transformation field, each of which would have to be searched in order to identify the relevant actual object.

Here are some elements of a more plausible account. First, recall that what is being rejected is an account according to which a perceptual system compares incoming perceptual data – assumed to be data that directly represents features of some relevant object X – with prototype features of various objects, until a match is found with features specifically of that object X. Instead, that initial search is performed on what is assumed to be contextual or aspectual *transformations* of data relating to some object X.

Now so far the matching problem is arguably intractable, because the data by hypothesis involves, or requires for its correct interpretation, *two*

kinds of factors – both the unknown object features, and whatever contextual factors are involved in the contextual transformation of those features – whereas the incoming perceptual data Y′(X′) only provides *one* of those factors, namely the resultant contextually transformed data. Thus the current account provides in effect a novel 'poverty of the stimulus' perceptual argument: there is not enough data in the stimulus itself to identify *both* of the relevant (at least partly) independent variables.

Nevertheless, when the problem is thus described, it becomes clear what is needed to resolve this specific 'stimulus poverty' problem, namely some *additional* perceptual information, that is *relatively independent* of the transformed data Y′(X′) from some particular object X that was involved in the relevant perceptual case. Or in other words, perceptual data identifying relevant contextual factors Y *as such* is required, in addition to the data Y′(X′) that specifically consists of transformed X-related data.

Once those contextual factors Y are known, it will then be possible for the perceptual system to calculate an *inverse transformation*, that will undo or *subtract* the effects of contextual factors Y on the transformed data Y′(X′), so as to calculate the initial features of *object X itself* that were transformed by those contextual factors Y.

For example, if a group of objects were being perceived in low light conditions, then all of the colors C of the objects would be systematically transformed in familiar ways, including a loss of brightness and color intensity. But the perceptual system could perceive that the light level is low, independently of perception of any *particular* transformed object in the group,[6] and hence *make allowances* for the general low light conditions, which allowances would involve the calculation of an inverse transformation, so as to arrive at the relevant actual color-related object features C for any given object in the group.[7]

In terms of the processing of perceptual information itself, any given set of perceptual data Y′(X′) that was being received, which includes transformed color data Y′(C′) for some object X, would be treated as providing, not the relevant color-related content C′ itself for object X, but instead as providing only a transformed or encoded version Y′(C′) of that content C′. So an inverse transformation would have to be performed by the system to calculate the relevant content element C′, which is hence perceptually *inferred* or *derived* rather than being perceptually immediate.

As for the initial concept of matching perceptual data with some element in a *transformation field* rather than directly with an object prototype, that is one useful way to conceptualize the relations between transformed data element Y′(X′) and X′ itself. The alternative idea of calculation of an inverse transformation from Y′(X′) to X′ is conceptually equivalent to

it, for in cognitive processing terms, any such relation may actually be implemented either by having a stored 'lookup table' of transformed field elements $Y'(X')$, with simple links to the corresponding object features X', or alternatively by not having such a table, but instead performing a real-time calculation of the relevant inverse function for each new item of transformed perceptual data $Y'(X')$, so as to produce the result X'. In general a stored transformation field would produce much faster identification, but at the cost of heavy field data storage requirements, while on the other hand, real-time calculations minimize storage requirements, but are much slower than the lookup table method. It seems likely that actual perceptual mechanisms would make use of each method as appropriate.

Thus, to summarize, the matching of perceptual data $Y'(X')$ with stored prototypes that include X-related information may be regarded as taking place either directly on elements of a transformation field, or instead as a more conventional matching process with an X-related prototype, that takes place *after* calculation of the relevant inverse perceptual data X' from the initial data $Y'(X')$.

5. THE RANGE OF PERCEPTUAL DOUBLE CONTENT CASES

My claim has been, not that all perception involves perceptual content that has a double content structure, but only that normal cases of perceptual identification and reidentification of real objects in a perceiver's environment, of a kind that would be subject to aspectual or contextual factors that would produce *contextually transformed* low-level perceptual data, would typically involve such a nested, two level structure of perceptual content.

Thus for example the perception of after-images, which do not involve real-world perceptual data of any kind, might involve only a single level of perceptual content. Also kinds of perception that involve only the perception of simple, unstructured events or states, such as the close-up seeing of a uniform expanse of color or the hearing of a continuous tone, might also be thought to involve only one level of perceptual content (but see the countervailing considerations below that favor uniform perceptual mechanisms for all cases).

However, more complex events such as the performance of a musical work, correct hearing of which would require such skills as the ability to identify and re-identify musical themes in various musical settings in the work, presumably would require a double content perceptual analysis, in that such works would have an aspectual structure of similar complexity to that possessed by persisting objects in natural environments.[8]

Returning to normal cases of perceptual identification, it might be thought that the double content account is only plausible for perception under *non-optimal* conditions. For if an object is seen under optimal conditions, including factors such as an object's being seen in good lighting, at close range, with no movement, and with its most salient side for accurate identification frontally presented to the viewer, would not any aspectual or contextual factors be reduced to zero, so that the transformed perceptual content $Y'(X')$ would be identical with the content X' for X itself?

However, that supposition would violate scientific plausibility considerations. To be sure, if *all* perception occurred under optimal conditions, it might be that perceptual content would have only a single level; but given that perceptual mechanisms have to identify objects under a wide range of conditions, most of which are non-optimal, evolutionary considerations virtually guarantee that only *general-purpose* mechanisms, capable of effectively identifying objects using the same uniform methods under *any* conditions, would survive. Given the *general* utility of perceptual mechanisms that result in perceptual double content, the mere fact that their use would occasionally result in some informational redundancy in optimal cases does not provide convincing evidence against the existence of such general-purpose, double content producing mechanisms.

Also, there is an interesting epistemological problem that would undermine the usefulness of a special-purpose, optimal-conditions perceptual mechanism for object identification, even if there were any such mechanisms. Optimal conditions for observation of an object depend on what kind of object it is; for example, seeing stars requires darkness, which is decidedly non-optimal for seeing most earthbound objects, for which bright light is optimal; but in turn those conditions are non-optimal for star observation. Hence in general, a perceiver cannot know what conditions are optimal for a given object until she has perceptually identified the object in question; but then it is already too late for her to make use of any special, optimal-conditions mechanisms during that identification.

The issue being discussed is related to that of the *transparency* of some representations. For example, some color photographs or trompe l'oeil paintings may be so realistic that perceiving them may phenomenologically be exactly like perceiving the actual scenes that they represent, with the perceiver having no distinctive awareness of the pictorial representational conditions in question; analogously, some perceptions of actual objects under optimal conditions may be so completely optimal that one has no distinctive awareness of the contextual conditions under which one is perceiving them. But the relevant perceptions in each case could still have a double content structure as a result of their causal origins from

standard perceptual mechanisms, even if their transparency made it hard to become introspectively aware of that structure.

6. INTERPRETATION IN PERCEPTION

This section provides an initial investigation of the very complex topic of the ways in which, or extent to which, various kinds of *interpretive* factors enter into perceptual processes – not as a detachable, indirect and secondary stage of *post*-perceptual processing, but instead as a direct and integral part of those perceptual processes themselves. Nevertheless, though preliminary, the investigations should serve to indicate at least the *potential* theoretical fruitfulness of a double content (DC) theory of perception, as applied to issues of perceptual interpretation.

Various kinds of interpretation seem to be involved in perception. One psychological kind involves interactions between the central *cognitive* aspects of perception, that are concerned with objective information gathering, versus various *conative* and *affective* aspects, such as motivation, desire, interests, attitudes and emotions, which are naturally regarded as causing a perceiver to *interpret* her initial, raw cognitive perceptual data in various ways. For example, a timid or fearful person in an unfamiliar part of town may wrongly interpret, or misperceive, a gesture of an inhabitant as being a *threatening* gesture, whereas the same gesture under more familiar circumstances would not have been so interpreted.[9]

The explanation to be given of such a case will be centered around the *context subtraction* process, by means of which the represented X-related content is calculated from the initial data $Y'(X')$ by subtraction or removal of the contextual information Y'. My suggestion is that any relevant conative or affective factors are primarily handled by the perceptual system as additional *external*, *aspectual* or *contextual* factors. In the example, the timid person's being in an unfamiliar part of town makes him perceive his *general situation*, including the environment he is in, as being a risky and intimidating one, *independently* of any particular persons or actions that he perceives. Thus his perceptual data involves both the normal cognitive element $Y'(X')$ and an affective/conative element Z', so that his total initial perceptual content is $Y'(X') + Z'$.

Here now is the crucial point. Conative and affective factors are not always completely understood by, or rationally under the control of, the person who is affected by them. Indeed, a fearful person may neither realize that he is feeling fear, nor be capable of rationally controlling it even if he did realize it. Hence, when the process of context subtraction is applied by the person to his perceptual data $Y'(X') + Z'$, he likely will *not*

adequately subtract or *properly discount* the additional conative/affective contextual factors Z′, in calculating the relevant X-related features that are represented by the aspectual information Y′(X′) + Z′.

Thus the end result is that the additional contextual factors Z′, or some part of them, end up *distorting* or *contaminating* the inferred or calculated X-related information, hence explaining how it is possible for the person to believe that he is genuinely *perceiving* a gesture by an inhabitant as being a threatening gesture – since the X-related perceptual content does indeed include some threatening-gesture content, as a result of the contextual contamination caused by the inadequate inversion process.

Contrast this account with a more conventional single-level account of perceptual content. On such a view, a person's perception of X and its properties is regarded as being processed independently of any contextual factors, so that a person's perceptual contents would presumably include both conventionally processed person-gesture information, resulting in harmless and non-threatening X-related perceptual content, and, as distinct items, the various anxiety-producing conative and affective factors. But on this account, there is no genuine *perception* of *specific* threatening gestures – not even any incorrect perception of such – and the view also inevitably predicts *conflicting* perceptual contents, with the harmless cognitive contents being in conflict with the potentially threatening conative and affective factors.

As a result, one would predict both *irresolution* on the part of persons thus perceiving a situation, because of the conflicting perceptual contents, and also an *uncertainty* as to whether a specific threat had genuinely been perceived, since a generalized fearful coloration of perception is presumably not specific or localized enough to count as a *specifically perceived* threat. However, typically persons who are not in rational control of various affective and conative factors have no such doubts or uncertainties about what they perceive, nor about what they should do, hence supporting the alternative double content view based on a context subtraction account of perceptual processing.

The conventional single-level account also threatens to leave any phenomenal aspects of the relevant conative and affective factors as dangling, intermediary qualia, having no clear representational role in perception, whereas the double content analysis instead offers support to a more unified, direct perceptual representationalism.[10]

Turning now to an important philosophical issue concerning perception and interpretation, the view that all perception, and hence observation in general, is *theory-laden* is a familiar one – that perceptual content never simply objectively represents objects or states of affairs as they ac-

tually are, but that instead it is inextricably intermixed with theoretical assumptions about, or interpretations of, what is being perceived, so that perception cannot provide genuinely independent observational evidence for a theory.[11]

On such a view, perception cannot be untangled from factors involving *theoretical interpretation* of low-level perceptual data: one's theories affect, not simply the indirect conclusions one later draws that are based on independent perceptual evidence, but more directly they affect the very content of that perceptual evidence itself.

A double content (DC) theory of perception can concede that a person's theories may in *some* way affect perceptual contents; for example, that a physicist observing a cloud chamber in an atomic collision experiment might in some sense actually *see* the collision and scattered particles, in the specific sense that she sees certain contextual *effects* produced by something in the cloud chamber, that she theoretically interprets as 'effects produced by colliding particles'. However, on the DC theory, such theory-related factors are in fact only part of the perceived *contextual aspects* that make up the initial, higher level of perceptual content; they do not also necessary affect or infect the lower, *represented* level of perceptual content – namely, of whatever it is that actually produces the theoretically interpreted effects – which perceptual content in turn externally represents the properties of the relevant particles and events. Thus one must distinguish the theoretically interpreted *contextual* perceptual content, from the object-related content nested within it.

As for any possible epistemic *bias* introduced by such a theoretical interpretation, all that the perceiver needs to do, in order for her to discount the potentially harmful effects of the theoretical contextual aspects, is for her to adequately *contextually subtract* them, in calculating or inferring the relevant non-contextual, object-related perceptual content. Or in other words, as long as the theoretical interpretive elements are perceptually used in an explicit and rational way – namely, one that conforms to adequate standards of scientific methodology – good scientists will be able to adjust their perceptual habits, involving inferences that produce object-related perceptual contents, so as to cleanse them of the explicit theoretical assumptions involved in the relevant higher-level perceptual aspects.

As an example, sound scientific methodology would require scientists to periodically view the perceptually observable results of experiments from *more than one* theoretical perspective. In so doing a scientist can train her own perceptual apparatus so as to produce an invariant lower-level perceptual content, no matter how theoretically different the higher-level perceptual contexts are in each case. Thus one may test for adequate

context subtraction, and remove any corresponding theoretical biases among experimenters, by this and other appropriate sound methodological practices.

7. PERCEPTUAL WAYS OF APPEARING

When one shuts one eye, and pushes on the sides of the other with one's fingers, a shifted, blurry image of the objects X in one's perceptual vicinity results. Presumably the resulting visual content is not a representation of blurred objects X, since the actual objects X are not blurred. But then how is the perceptual content to be explained? Any theory of perception must be able to give some adequate account of such cases.

One common kind of explanation would invoke qualia: the blurry objects in one's visual field are phenomenal entities that cannot be explained in representational terms.[12]

However, the double content (DC) theory of perception has a more economical explanation of such cases, that can avoid the postulation of qualia and hence potentially favor the representationalist position. It is that the blurry aspects Y′ of the perceptual content are aspectual or *contextual* factors in one's perceptual content Y′(X′), so that the relevant perceptual state S represents both a blurred aspectual state Y(X) associated with the objects – an unusual state that is generally of no scientific or practical interest – and the objects X themselves, which are of course unblurred. Thus the content of perception Y′(X′) can both be blurry, and yet be the content of a perceptual state that represents unblurred objects, because there are *two* levels of perceptual content, that together can satisfactorily account for representation both of the relevant object-related aspects, and of the objects themselves.[13]

Another more intuitive way to conceptualize the situation is in terms of a distinction between 'the content' of the perception, namely the object-related content, versus the *way in which* that content is perceived by the person in question, or in other words its *perceptual mode of presentation* to her. Being presented in a *blurry* way or manner is of a piece with being presented at a distance, in fog or mist, while moving, and so on: all involve contextual conditions under which the perceiving occurs, and so all of them can be explained in a similar way by the DC theory, without having to postulate qualia or other irreducibly non-representational 'ways of appearing'.

More generally, the equally intuitive idea that perception primarily gives information about the *appearances* of things, rather than about how they *actually* are, is also explained by the DC theory, in that the concept of

an appearance is simply a generalized phenomenal version of the concept of aspects or contextual conditions under which perception occurs. Indeed, on the DC view the higher-level aspectual perceptual data has a kind of *epistemic priority*, in that it provides the raw actual data, from which the nested lower-level object-related data is calculated by inversion or subtraction techniques. Thus only that higher-level data can provide genuinely new *perceptual information* to the perceiver, since the lower-level, object-related data must be inferred from it by the perceiver.[14]

Another more specific case that can be illuminated by the DC approach is as follows. One powerful recent argument against representationalism (the view that the phenomenal character of experiences is determined by their contents) is that visual versus auditory experiences differ in phenomenal character, even in so far as they represent similar properties of a given object.[15]

However, on the present view, the phenomenal differences can be located in the relevant higher-level contextual or aspectual perceptual contents, which are closely associated with aspectual representational differences. *Auditory* perceptual states represent both physical auditory aspects – such as sound-wave refractions – as caused by an object's properties, and also those properties themselves, while *visual* perceptual states correspondingly represent both physical visual aspects – such as light refractions – as caused by the object's properties, and also those properties themselves. Thus a DC theory of perception allows the distinctive features of each sensory modality to play a role in perceptual content, without swamping or distorting the common lower-level content in each case, hence supporting representationalism even in such difficult intermodal cases.

The DC approach could also be used to defuse arguments to the effect that there are no genuine perceptual contents such as experiences of the redness of an object, but that instead there is just a series of various perceptual interactions with the relevant object.[16] O'Regan and Noë argue that there is no genuine red perceptual content in perception of red objects, because the variety of the relevant perceptual interactions precludes any such common element. However, on the DC theory, the existence of such a variety of aspectual perceptual interactions is entirely compatible with perceptual states also representing a property of redness. Thus even if the authors are correct in rejecting a static, aspect-free representational view of perception, the current DC theory could still accommodate their findings.

To conclude this section, another very important concept that a DC theory might be usable to explain is that of *what it is like* to be a perceiver situated in a specific perceptual situation, in that the theory is built around

a *perspectival*, contextual view of incoming perceptual information. Then the more abstract or metaphysical concept of what it is like to be a *human being*, rather than, say, a bat,[17] may be at least partly explicated in terms of the characteristic contextual features of human perceptual mechanisms when used under typical perceptual conditions, as opposed to those of bats under their very different perceptual conditions.

8. CONCLUSION

In conclusion, here is a brief discussion of questions concerning the empirical basis of the current theory, including the relation of the theory to scientific studies of perception.

First, the account given of the structure and genesis of perceptual information is at a very general level of description, and hence it is not intended to be a specific hypothesis about the actual workings of any given stage of perceptual processing.

The main empirical evidence for the DC theory is centered around the contrast between two categorial kinds of perceptual content – of aspects of objects, versus the objects themselves. Insofar as conscious perceptual experience does involve both kinds of content, *some* account must be given of how that two level, double content structure was arrived at. Thus my hypothetical reconstruction, according to which perceptual processes involve – in some way, or at some level or levels – a procedure of *context subtraction* as applied to initially complex, contextually loaded perceptual data, does provide one plausible basic account that would explain the resultant content structure.

However, an alternative procedure for arriving at the same double content result might also play some part in at least some actual perceptual mechanisms. Suppose that there are some initial, sub-personal object identification processes in which object features are identified *directly*, with contextual elements playing no significant initial role.[18]

For such cases, if any, an alternate process of *context addition* could be postulated, in order to explain the resultant double content structure of perception. Thus the initial context-independent object identifications would be supplemented by appropriate contextual additions, so as to explain how perceptual contents include both *objects*, and yet also specific contextual *aspects* of those objects, that are experienced as being *of* those objects. It will be work for future papers to investigate actual cases, so as to correctly classify them as contextual subtraction or contextual addition cases – and in addition, to provide more detailed accounts of the rest of the significant issues discussed here.

ACKNOWLEDGEMENT

My thanks to an anonymous reviewer for very helpful comments on previous versions of this paper.

NOTES

[1] As opposed to common 'poverty of the stimulus' arguments in cognitive science.

[2] This is a much broader issue than that of perceptual constancy. It covers not just cases such as those in which an object continues to look white under different illumination conditions, but also cases where aspects of an object do look different to each other, but nevertheless are still perceived as aspects of the same object.

[3] Linguistic and intentional higher order representations (metarepresentations) are very actively studied currently – see, e.g., Sperber (2000). But pictorial and perceptual cases of the relevant second order kinds have been neglected.

[4] Hence it is also premature to inquire as to the precise relations – beyond the overt differences – between the current concept of nesting and Dretske's non-semantic, purely information-theoretic concept of nesting as presented in his book *Knowledge and the Flow of Information* (1981).

[5] This is a very simplified account. For an overview of perceptual theories that include such elements see Biederman (1995).

[6] Some independent knowledge might also play a part in some cases, such as the perceiver's knowing that it is evening, or that the sun has set.

[7] To be sure, such an operation would give only an approximate result in the case of low light or other non-optimal conditions, but such perceptual imperfections are a fact that any analysis of perception has to come to terms with.

[8] A general representational theory of aesthetics is being developed along these lines; see, e.g., Dilworth (2001), (2003) and also my forthcoming book *The Double Content of Art*.

[9] For a general survey of some relevant issues, see Niedenthal and Kitayama (1994).

[10] Of the sort proposed by Dretske, Tye, Lycan et al. A generalized version of this direct view is described as 'intentionalism' by Alex Byrne in his (2001).

[11] E.g., see Kuhn (1962).

[12] For arguments against such a view see, e.g., Tye (1997).

[13] To be sure, other arguments for qualia, including inverted/twin earth arguments and Jackson's knowledge argument, are not addressed here.

[14] As for how, on this account, an organism is ever able to learn object-related information in the first place, the answer is, as implied by the discussion in Section 4, that under optimal conditions only trivial inverse transformations are needed to calculate such object-related information. See also Section 8 on how other perceptual mechanisms may involve 'context addition' methods that would permit direct learning of object-related information.

[15] See Lopes (2000).

[16] E.g., see O'Regan and Noë (2001).

[17] See Nagel (1974).

[18] As in Biederman's geon theory, e.g., see Biederman (1987).

REFERENCES

Biederman, I.: 1987, 'Recognition-by-Components: A Theory of Human Image Understanding', *Psychological Review* **94**, 115–147.

Biederman, I.: 1995, 'Visual object recognition', in S. F. Kosslyn and D. N. Osherson (eds.), *An Invitation to Cognitive Science*, Vol. 2, *Visual Cognition*, Chap. 4, MIT Press, Cambridge, MA, pp. 121–165.

Byrne, A.: 2001, 'Intentionalism Defended', *The Philosophical Review* **110**, 199–240.

Dilworth, J.: 2001, 'A Representational Theory of Artefacts and Artworks', *The British Journal of Aesthetics* **41**, 353–370.

Dilworth, J.: 2003, 'Medium, Subject Matter and Representation', *The Southern Journal of Philosophy* **41**, 45–62.

Dretske, F.: 1981, *Knowledge and the Flow of Information*, MIT Press, Cambridge, MA.

Kuhn, T.: 1962, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

Lopes, D. M. M.: 2000, 'What Is It Like to See With Your Ears? The Representational Theory of Mind', *Philosophy and Phenomenological Research* **60**, 439–453.

Nagel, T.: 1974, 'What Is It Like to Be a Bat?', *The Philosophical Review* **LXXXIII**, 435–450.

Niedenthal, P.M. and S. Kitayama (eds.): 1994, *The Heart's Eye: Emotional Influences in Perception and Attention*, Academic Press, New York.

O'Regan, J. K. and A. Noë: 2001, 'A Sensorimotor Account of Vision and Visual Consciousness', *Behavioral and Brain Sciences* **24**, 939–973.

Sperber, D., ed.: 2000, *Metarepresentations*, Oxford University Press, Oxford.

Tye, M.: 1997, 'A Representational Theory of Pains and Their Phenomenal Character', in N. Block, O. Flanagan and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates*, MIT Press, Cambridge, MA.

Department of Philosophy
Western Michigan University
3004 Moore Hall
Kalamazoo, MI 49008
U.S.A.
E-mail: dilworth@wmich.edu